DOCUMENT RESUME

ED 164 599 95 TM 008 148

AUTHOR Guba, Egon G.
TITLE Toward a Methodology of Naturalistic Inquiry in Educational Evaluation. CSE Monograph Series in Evaluation, 8.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE 78
NOTE 97p.; Small print may be marginally legible
AVAILABLE FROM Center for the Study of Evaluation, Graduate School of Education, University of California, Los Angeles, California 90024 ($4.50)

EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS *Credibility; *Educational Assessment; Evaluation Criteria; *Evaluation Methods; *Observation; Program Evaluation; Reliability; Research Design; *Research Methodology; Scientific Methodology; *Validity
IDENTIFIERS Evaluation Problems; *Naturalistic Observation

ABSTRACT

Evaluation is viewed as essential to decision making and social policy development. Since conventional methods have been disappointing or inadequate, naturalistic inquiry (N/I) differs from conventional science in minimizing constraints on antecedent conditions (controls) and on output (dependent variables). N/I is phenomenological rather than positivist. It offers alternative strategies for problems when the experimental approach is implausible. A number of new evaluation models (such as the responsive model, the judicial model, and the connoisseurship model) are compatible with the approach. Since there is no compelling way to truth, N/I must be credible and deal convincingly with standard methodological problems such as boundary problems (setting the scope of inquiry), focussing problems (establishing and defining categories), and problems of authenticity (reliability, validity, and objectivity). Techniques for establishing validity include: triangulation, cross examination, persistent observation, and peer or participant corroboration. In evaluation, validity may be ecological, contextual, or phenomenological. Impartiality is imperilled by conscious or unconscious bias, incompetence, gullibility, or corruptibility. It is promoted by openness and fairness. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

CSE
MONOGRAPH
SERIES
IN
EVALUATION

8

ED164599

TM008 148

TOWARD A METHODOLOGY OF NATURALISTIC INQUIRY IN EDUCATIONAL EVALUATION

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Egon G. Guba

Carol C. Kiser

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

CENTER FOR THE STUDY OF EVALUATION
UNIVERSITY OF CALIFORNIA • LOS ANGELES

2

CSE MONOGRAPH SERIES
IN EVALUATION

SERIES EDITOR

Eva L. Baker

Center for the Study of Evaluation
UCLA Graduate School of Education
University of California, Los Angeles
Los Angeles, California 90024

3

TOWARD A METHODOLOGY OF NATURALISTIC INQUIRY IN EDUCATIONAL EVALUATION

Egon G. Guba

Center for the Study of Evaluation
UCLA Graduate School of Education
University of California, Los Angeles, 1978

CSE MONOGRAPH SERIES IN EVALUATION

NUMBER

1. Domain-Referenced Curriculum Evaluation: A Technical Handbook and a Case Study from the MINNEMAST Project
   Wells Hively, Graham Maxwell, George Rabehl, Donald Sension, and Stephen Lundin — $3.50
2. National Priorities for Elementary Education
   Ralph Hoepfner, Paul A. Bradley, and William J. Doherty — $3.50
3. Problems in Criterion-Referenced Measurement
   Chester W. Harris, Marvin C. Alkin, and W. James Popham (Editors) — $3.50
4. Evaluation and Decision Making: The Title VII Experience
   Marvin C. Alkin, Jacqueline Kosecoff, Carol Fitz-Gibbon, and Richard Seligman — $3.50
5. Evaluation Study of the California State Preschool Program
   Ralph Hoepfner and Arlene Fink — $3.50
6. Achievement Test Items–Methods of Study
   Chester W. Harris, Andrea Pastorok Pearlman, and Rand R. Wilcox — $4.50
7. The Logic of Evaluative Argument
   Ernest R. House — $4.50
8. Toward a Methodology of Naturalistic Inquiry in Educational Evaluation
   Egon G. Guba — $4.50

This Project was supported in whole or in part by the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

TABLE OF CONTENTS

Foreword vii

Chapter I: Introduction 1

Chapter II: The Nature of Naturalistic Inquiry (N/I) 3

- Some Definitions 3
- How Naturalistic Inquiry Differs from More Conventional Inquiry 11
- How Naturalistic Inquiry Is Similar to Other Methodologies 18
- The Need for Naturalistic Inquiry 23

Chapter III: The Utility of Naturalistic Inquiry for Educational Evaluation 31

- Theoretical Movement Away from Conventional Paradigmatic Models 31
- Practitioner Movement Away from Conventional Evaluation Methodologies 32
- Emergent Evaluation Models Congenial to the Use of Naturalistic Inquiry as a Supporting Tool 33
- The Fit Between Evaluation's Main Thrusts and the Strengths of Naturalistic Inquiry 41

Chapter IV: Selected Methodological Problems of Naturalistic Inquiry 42

- Boundary Problems 44
- Focussing Problems 49
- Problems of Authenticity 61

Chapter V: Summary 79

Bibliography 82

LIST OF FIGURES

FIGURE	Page
1 A "Wave" Conceptualization of Naturalistic Inquiry	7
2 Representation of Inquiry Space or Domain of Inquiry	10
3 Representation of Divergence Strategies: Extending, Bridging, and Surfacing	58
4 Representation of Quantitative and Qualitative Aspects of Objectivity/Subjectivity	74

FOREWORD

The mission of the Center for the Study of Evaluation (CSE) is to conduct inquiry, from a variety of perspectives, into the nature of evaluation as it affects the formulation, development, and delivery of educational programs and services. Our commitment to inquiry into the field of evaluation grows from the belief that school practices and the competencies and satisfactions of those who participate in the educational enterprise can benefit from information collected in accordance with social science methodology. At CSE we study the instruments and methodologies for collecting and interpreting information as well as the sociopolitical context of educational decision making.

Evaluation as a field came into existence in part as a response to the political problem of insuring accountability for federally funded programs, and in part as a response to public demands for more information about how schools are or are not serving the educational needs of children. Growth in the field has been marked by an increase in its size--that is, by ever greater numbers of people calling themselves evaluators--and by an increase in federal, state and private expenditures. Growth within the field has also been marked by evaluators putting away childish things--among them, simple-minded views of the world, fantasies of perfection and of control, reliance on ready-made solution to non-standard problems.

Evaluation may be regarded as in its adolescence, with one physical and intellectual growth spurt behind it and more perhaps still to come. This adolescence is characterized by concern with identity: What are the purposes and who are the audiences that evaluators serve? By concern with standards: What is a "good" evaluation? By concern with ethics: How do evaluators resolve the tension between client responsiveness and professional objectivity? And with ways of knowing: What are the methodologies available to evaluators, and under what circumstances are they appropriate?

It is to this last issue that Dr. Egon Guba has addressed himself. In his monograph, Dr. Guba makes the case for naturalistic inquiry as distinguished from experimental inquiry. He then examines three major conceptual problems in doing evaluations using naturalistic methodology: boundary problems, focusing problems, and problems of authenticity. In so doing he summarizes work from other fields and demonstrates their applicability to evaluation. He also provides a stimulus to new thinking for others who are now using such methodology. Although he doesn't describe specific procedures or list the skills needed to train competent naturalistic investigators, his framework supports those who would take these next steps.

Dr. Guba was a resident Visiting Scholar at CSE during the summer of 1977. The Visiting Scholars program brings to the Center eminent individuals

who have a substantial contribution to make to the theory and practice of evaluation. This monograph is one visible outcome of the Visiting Scholars program. Other less tangible but equally valuable benefits derive from Dr. Guba's participation in another CSE program—that dealing with Evaluation and Decision Making. Here CSE studies the state of the art in evaluation from three perspectives: organizational, interpersonal and philosophic.

We anticipate that CSE's encouragement of the study of evaluation will arouse interest and discussion. We welcome your comments.

Adrianne Bank
Associate Director

ACKNOWLEDGEMENTS

The first draft of this monograph was written during my residence as a Visiting Scholar at the Center for the Study of Evaluation, Graduate School of Education, University of California, Los Angeles, during the summer of 1977. I am indebted to the Director, Eva Baker, for having made this opportunity available to me, and to the Associate Director, Adrianne Bank, for her continuing support and encouragement.

The original stimulus for this volume came from three colleagues, Yvonna Lincoln, Assistant Professor of Higher Education at the University of Kansas; Barbara Tymitz, Research Associate in the Center for Innovation in Teaching of the Handicapped, Indiana University; and my associate in the Educational Inquiry Methodology program at Indiana University. Robert Wolf, an outstanding scholar and practitioner in the area covered by this essay. I am grateful to them for urging me to undertake this monograph and for providing valuable counsel and review at various stages of its writing.

Several readers of the early draft were kind enough to offer useful comments, including Harbans Bhola, Fred Ellett, Arthur Lumsdaine, Lynn Morris, W. James Popham, Richard Pugh, David Rindskopf, Robert Stake, and Louise Tyler. I have profited from their input but of course must retain sole responsibility for any errors of fact or interpretation herein contained.

Finally I must acknowledge a debt to Shelby Popham, who spent long hours editing the manuscript, and to Jayne Steele who did the final typing.

Egon G. Guba

School of Education
Indiana University
Bloomington, Indiana
February 1, 1978

Truth is not that which is demonstrable; Truth is that which is ineluctable.

–Antoine de St. Exupery

The competition between paradigms is not the sort of battle that can be resolved by proofs. . . . The transfer of allegiance from paradigm to paradigm is a conversion experience that cannot be forced.

–Thomas S. Kuhn

Although I am fully convinced of the truth of the views given in this volume, . . . I by no means expect to convince experienced naturalists whose minds are stocked with a multitude of facts all viewed, during a long course of years, from a point of view directly opposite to mine. . . . But I look with confidence to the future, to young and rising naturalists, who will be able to view both sides of the question with impartiality.

–Charles Darwin

Chapter I:

INTRODUCTION

Evaluation has come to be widely viewed as the handmaiden of decision-making and social policy development. Congress has mandated that evaluation shall be an integral part of almost every Federally supported social action program--whether in education, health, justice, or welfare--both to assist in the development and refinement of these programs as well as to render some judgment about their worth. Many state agencies, school systems, hospitals, courts, and municipalities now support either evaluation offices or staffs of individuals charged with the evaluation function. Information management systems, many of which display characteristics of evaluation units, are also common.

Yet many agree that evaluation has not been as fruitful as had been hoped or expected. Useful evaluation information is not often produced; and even when it is, decision-makers and policy formulators sometimes see fit to disregard it. While there are undoubtedly many reasons for this state of affairs, a major reason seems to be the lack of a methodology uniquely suited to evaluation's needs.

One alternative to conventional evaluation methodology is that of *naturalistic inquiry*,[1] an approach which has considerable promise for social and behavioral inquiry generally and for evaluation particularly. Although naturalistic inquiry is not a new method, having its roots in ethnography[2] and phenomenology, it has typically been eschewed as a legitimate method because of its "softness." Thus, Rossi and Wright (1977) could say, in the lead issue of the new journal *Evaluation Quarterly*, that

> There is almost universal agreement among evaluation researchers that the randomized controlled experiment is the ideal model for evaluating the effectiveness of public policy. If there is a Bible for evaluation, the Scriptures have been written by Campbell and Stanley. (p. 13)

But Wolf and Tymitz (1977) take sharp exception to this judgment:

> Some people assume that natural inquiry is unstructured, undisciplined, and unsystematic; let it suffice to say that such thoughts are unfair, inaccurate, and simplistic. Natural inquiry is no less rigorous than traditional experimental research, it is simply different. The paradigm of natural inquiry is comprehensive in scope, demanding in design, and

[1] The term "naturalistic inquiry" will frequently be abbreviated "N/I."

[2] The term *ethnography*, indicating that branch of anthropology dealing descriptively with cultures, was first used by Malinowski in 1922. See Malinowski (1922).

> requires a set of honorable skills that even some rigid experimentalists lack (although they might be embarrassed to admit it). (p. 7)

Denzin (1971) echoes their sentiments by pointing out that

> the naturalist is committed to *sophisticated rigor*. Which is to say he is committed to making his data and explanatory schemes as public and replicable as possible. Thus he details in careful fashion the nature of his sampling framework, triangulates his observations and continually assesses the empirical grounding of his causal propositions. (p. 167)

It is a truism among epistemologists that there is no way of compelling intellectual concurrence about any knowledge. Formulations about truth always depend heavily on a host of axiomatic propositions which cannot really be put to the test. To protect themselves against charlatans and to provide a basis for their own beliefs, persons seeking to "know" have evolved the idea of "method," and have defined "truth" as the result of an appropriate inquiry methodology. The experimental method has great force precisely because inquirers have agreed among themselves that they will trust (concur in, regard as acceptable, be persuaded by) the results of experimental inquiry. *If any other method is to be taken seriously as an alternative, bases for a similar trust will have to be convincingly argued.*

Thus the purpose of this essay is to delineate and, so far as possible, deal with middle range methodological questions relating to one such proposed alternative, naturalistic inquiry. The term "middle range" is used with care, to indicate issues that fall somewhere between the poles of epistemology and technique. This paper will not systematically take up questions pertaining to phenomenology as a philosophical position nor those detailing how naturalistic inquiry can in fact be operationalized (questions of technique). Both these areas will be touched upon, but for the sake of background more than for the sake of exposition. Works touching on the former are widespread; for an excellent discussion of the latter the reader is referred to the manual developed by Tymitz and Wolf (1977).

The remainder of this monograph discusses several issues in regard to naturalistic inquiry. Chapter 2 takes up the question of what naturalistic inquiry is and how it differs from more conventional forms of inquiry. Chapter 3 treats the utility of naturalistic inquiry for educational evaluation. Three major methodological problems that must be dealt with by proponents of naturalistic inquiry are described in Chapter 4: *boundary* problems, *focussing* problems, and *authenticity* problems. The monograph ends with a summary and a bibliography.

Chapter II:
THE NATURE OF NATURALISTIC INQUIRY (N/I)

Naturalistic inquiry (N/I) is an alternative mode of inquiry which differs from other modes by its relative position along two dimensions: (a) the degree of manipulation of conditions antecedent to the inquiry, and (b) the degree of constraint imposed on outputs by subjects involved in the inquiry. It also differs from other modes with respect to its basic assumptions or characteristics. Both these ideas will be discussed in some detail in ensuing paragraphs, but before turning to that discussion it may be instructive to review some definitions of naturalistic inquiry offered by other students of that process.

SOME DEFINITIONS

Naturalistic inquiry is defined in a variety of ways. There seems to be no systematic definition agreed upon by everyone. Educationists seem somewhat less formal about the definitions they offer. House (1977) for example, says,

> I would label as "naturalistic" evaluation that evaluation which attempts to arrive at naturalistic generalizations on the part of the audience; which is aimed at non-technical audiences like teachers or the public at large; which uses ordinary language; which is based on informal everyday reasoning; and which makes extensive use of arguments which attempt to establish the structure of reality.

While this definition refers specifically to naturalistic *evaluation*, it does outline many of the elements which seem to characterize N/I more broadly. However, the definition could hardly be labelled as systematic or rigorous.

Similarly, Wolf and Tymitz (1976-77) suggest that N/I is an inquiry mode aimed at understanding

> actualities, social realities, and human perceptions that exist untainted by the obtrusiveness of formal measurement or preconceived questions. It is a process geared to the uncovering of many idiosyncratic but nonetheless important stories told by real people, about real events, in real and natural ways. The more general the provocation, the more these stories will reflect what respondents view as the salient issues, the meaningful evidence, and the appropriate inferences. ... naturalistic inquiry attempts to present "slice-of-life" episodes documented through natural language and representing as closely as possible how people feel, what they know, how they know it, and what their concerns beliefs, perceptions, and understandings are.

This definition seems to focus on people as subjects for N/I and interactions with people, probably through interviews, as the typical N/I data collection method. The emphasis on "slice-of-life" episodes suggests the informality of the approach as viewed by these authors.

Probably the most systematic attempt to formulate a definition of N/I was made by Edwin P. Willems and Harold L. Raush (1969) in a collection of papers edited by them under the title, *Naturalistic Viewpoints in Psychological Research*. The volume presents ten papers ranging over fields as disparate as primate behavior, third grade classrooms, and ecological psychology. Willems and Raush themselves, in an introductory statement, provide a general purpose definition of N/I as "the investigation of phenomena within and in relation to their naturally occurring contexts" (p.3). In a summary chapter entitled, "Interpretations and Impressions," however, these two authors analyze definitions offered by several chapter contributors and attempt a synthesis. A review of this material seems in order.

E. W. Menzel, a primate anthropologist, sees N/I as the broadest form of inquiry:

> There is . . . no logical basis for the current attitude . . . that naturalistic and experimental research are in opposition to each other. This can hardly be so, because naturalistic research is the more generic. Naturalistic research is any form of research that aims at discovery and verification through observation, and this includes as specialized cases all techniques, apparatus, and procedures of experimentation. (Willems & Raush, 1969, p. 81)

Menzel provides an interesting analog for his suggestion that naturalistic and experimental inquiry are in fact related:

> Here the field worker and the laboratory man, insofar as they are scientific organizers of data rather than catalogers or technicians, tend to adopt different but compatible methods of achieving perspective. The methods are analogous to *zooming in* and *zooming out* with a lens. To the extent that they are reproduced objectively, wide-angle, telephoto, and microscopic views must be *simultaneously* valid, and zooming from different directions merely focusses attention on different facets of the same phenomenon . . . There are no grounds, logical or otherwise, for calling any view "simple." We can start anywhere and zoom in to infinite detail, or zoom out to infinite scope. (Willems & Raush, 1969, pp. 82-83)

Lee Sechrest, a psychologist, in discussing the assessment of attitudes suggests that

> There are any number of ways in which an attitude may manifest itself, and, with one exception (autonomic response measures), the manifestations are not limited to the laboratory. It is the *stimulus situation*, not the response, that defines a naturalistic method (emphasis added). (Willems & Raush, 1969, p. 152)

Reasoning along these lines, Sechrest concludes that

> What are needed in the study of social attitudes, as in other areas of psychology, are measures which (a) do not require the cooperation of the subject, (b) do not permit the subject's awareness that he is being measured or treated in any special way, and (c) do not change the phenomenon being measured. The above three provisions may be taken as an operational definition of *naturalistic* measures. (Willems & Raush, 1969, p. 152)

Roger Barker, a psychologist widely known as the "father" of ecological psychology, has called for the development of an "eco-behavioral" science:

> The eco-behavioral science that will answer the pressing questions society faces today requires, above all, concepts and theories appropriate to the phenomena involved. But these will not arise *de novo*; they will be grounded upon empirical data concerning the patterns of events within the psychologist-free settings where people live their lives. Special facilities are required in order to obtain these data.
>
> . . . In ecological investigations, where phenomena are studies *in situ*, nature is the only inducer and the investigator can be only a transducer . . . (Willems & Raush, 1969, p. 37)

David Gutmann, a psychologist particularly interested in cross-cultural studies, makes an interesting distinction between the "naturalist" and the "theorist":

> For me, the crux of the naturalist's method is that it does not treat nature as passive. The naturalistic assumption in any field, is that intrinsic orders exist 'out there' and that these regularities will organize and drive events even though our theories take no notice of them. By contrast, the experimentalist's implicit assumption is that nature is passive; that it exists because theories exist, and that the role of the environment is to supply data that will either demonstrate theory or refute it. . . . Thus, the techniques and instruments of the naturalist are aimed at bringing out, at highlighting some implicit order in the domain of his interest, and toward turning the implicit order into explicit data. Where the task of the theorist is to explain data, the special task of the naturalist is to generate data. (Willems & Raush, 1969, p. 162)

Harold L. Raush, a clinical psychologist, makes a somewhat more operational distinction between the experiment and the field study:

> Consider the experiment in contrast to the field study. In the laboratory experiment, the investigator defines and restricts the ranges of both stimuli and responses. For example, in a narrow case, the stimuli may be a red and a green light, and the response the pressing of one button or another. . . . A polar contrast is the purely naturalistic study in which the investigator attempts to exercise no influence over the range of stimuli or the range of responses; it is the subject who selects and defines the repertoire of both stimuli and responses. (Willems & Raush, 1969, p. 130)

Thus a study can be characterized as experimental or naturalistic as a function of where it falls along two dimensions of manipulation by the investigator: the manipulation of stimuli, and the manipulation of response modes, or at least, the designation of which responses will be noted.

This very important distinction is made more explicit by Edwin P. Willems, a psychologist:

> In behavioral research, naturalness or naturalism . . . is a function of *what the investigator does*. . . . the set of activities an investigator actually engages in while conducting his research falls somewhere in a two-dimensional descriptive space. The first dimension, which is most frequently thought of in differentiating research activities, describes the degree of the investigator's influence upon, or manipulation of, the antecedent conditions of the behavior studied, on the assumption that the degree of such influence or manipulation may vary from high to low, or from much to none. The second dimension, which is less commonly considered than the first, describes *the degree to which units are imposed by the investigator upon the behavior studied.* (Willems & Raush, 1969 p.46)

Willems goes on to say that experimental inquiry would fall into the high-high portion of the field described by these two orthogonal dimensions, while naturalistic inquiry would fall into the low-low portion.

Willems and Raush, after briefly summarizing the proposals made by the six authors cited above, go on to identify six "converging" descriptors or characteristics of naturalistic inquiry that may be deduced from them (with varying degrees of support for each):[3]

1. *Naturalistic inquiry is always a matter of degree.* This generalization is particularly evident if one considers the proposals, individually, by Willems and Raush that inquiry can be characterized as a function of the degree of manipulation of stimulus and response modes. N/I cannot be "pure" in the sense of being absolutely free of constraints placed on either antecedents or responses; initial efforts by the investigator to discover the meaning of what he has observed will cause him to propose certain categories in which to assimilate and account for the noted responses. These categories lead him further into a verification mode, so that on subsequent observations he is more likely to select situations that elicit the response categories of interest. Most likely, the investigator will *cycle* through a series of observations that are, alternately, directed at discovery and then at verification; some initial verification leads to a reorientation to further discovery, and so on.

This phenomenon can be conceptualized as a kind of "wave"[4] of the sort depicted in Figure 1. At the beginning of the inquiry (time = t_0) the natural-

[3]The discussion which follows was suggested by the treatment provided in Willems and Raush, 1969, p. 273.

[4]I am indebted to Barbara Tymitz for this concept.

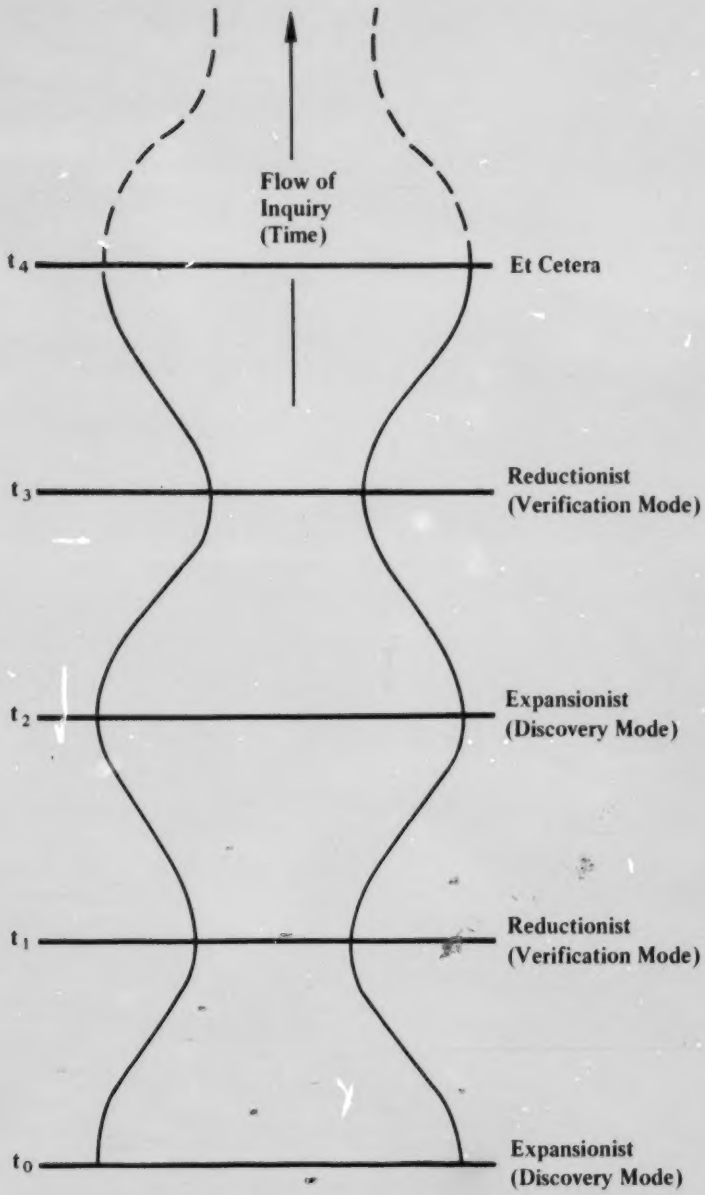

Figure 1: A "Wave" Conceptualization of Naturalistic Inquiry

istic investigator adopts the discovery mode, which might also be termed the "expansionist" mode. After some period has passed, at time t_1, some hypotheses (perhaps mere "hunches") may have emerged prompting the investigator to adopt the verification mode, which might also be termed the "reductionist" mode. Later, at time t_2, the investigator is moved once more to open his inquiry to new inputs; he slips back into the discovery/expansionist mode. Similar alternations characterize subsequent times (t_3, t_4, . . .). Inquiries, even those that are avowedly naturalistic, are thus almost always comprised of *mixed* strategies.[5]

2. *The degree to which a study is naturalistic is a function of what the investigator does.* It is important to note that naturalistic inquiry is not determined by the nature or posture of the subjects or of the situation, but rather by what the investigator chooses to do, i.e., the particular methodology which he chooses to apply. Moreover, this choice is a relatively open one for the investigator, i.e., there is nothing intrinsic to particular areas of inquiry that forces the investigator to adopt one or another of the several modes of inquiry.

3. *What the investigator does in relation to stimuli, independent variables, or antecedent conditions is one crucial dimension.* This dimension is the first of the two referred to by both Willems and Raush, and implied in all of the six other authors' formulations. Manipulation of any element prior to observation is, for all authors, sufficient reason to regard a study as *not* naturalistic.

4. *What the investigator does in restricting the response range or domain of the subject's output is a secondary dimension in the minds of two of the authors (Willems & Raush).* Manipulation of output possibilities is, in their opinion, also sufficient to label a study as *not* naturalistic.

5. *Three of the authors (Barker, Menzel, & Gutmann) suggest that naturalistic inquiry provides the means for an investigator to approach behavioral phenomena as if for the first time, with minimal determination by prior theoretical categories.* Naturalistic inquiry does not oblige the investigator to have formed certain conceptions or theories about his field of interest; instead he can approach it with pristine mind and allow his interpretations to emanate from and be influenced by real events, rather than the reverse. But, it should be noted, the remaining three authors do not share this position. Moreover, the "wave" concept explicated above also suggests that insistence on a conceptually blank approach may be both inappropriate and naive.

6. *The term "naturalistic" is understood by all authors to be a term that modifies "research" or "method" but not "phenomena."* A warning, "Beware

[5]But the reader should note that the verification phases are also carried out in a naturalistic "style." See p. 14 for a characterization of that style.

of Greeks claiming naturalness" may be appropriate here. It is hard to find an idea in psychology, sociology, or any of the behavioral sciences that could not somehow be advanced on the ground that it was more "natural" than its predecessors. (In education, this claim is very frequently made for new teaching methods.) As Willems and Raush put it,

> the authors do not seem to find it useful to speculate about how natural the *findings* of a study are, but they do find it worthwhile to differentiate the *activities of investigators*. Thus, while we may all have intuitive ideas as to what natural events are, the concept of "natural" here becomes translated into an arbitrary label attached to investigative activities that fall at one end of a complex spectrum. (p. 273)

Despite this analysis of "converging" descriptors or characteristics of naturalistic inquiry, one is still left with the feeling that there are many ways to view N/I. Nevertheless, it would probably be useful to focus on one of them here for the sake of clarity and precision. The option with seemingly highest appeal is the formulation of Willems which suggests that there are two dimensions defining the domain of inquiry: the extent to which the investigator places constraints upon antecedent conditions, on the one hand, and possible outputs, on the other. Willems has developed a chart which displays these two dimensions orthogonally and permits him to label various portions of the field appropriately; e.g., as "experimental," "naturalistic," and so on (Willems & Raush, 1969, p.47). Figure 2 is an adaptation of his chart; however, for simplicity's sake the two continua have been reduced to "low" and "high" segments. The domain of inquiry is thus represented as a 2X2 table.

Experimental inquiries fit into the upper right hand cell, in which constraints on both antecedent conditions and possible outputs are sharply defined. In this cell, the investigator decides both upon the nature of the independent variables and how they shall be manipulated (e.g., subjects' IQs shall be divided into three ranges for low, medium, and high intelligence), and upon the nature of permissible outputs (e.g., number of nonsense syllables that can be recalled immediately and ten minutes following memorization).

Naturalistic inquiries fit into the lower left hand cell, in which neither dependent nor independent variables are manipulated by the investigator. We have already noted the likely "wave" nature of these inquiries, but it seems probable that the full "wave," as depicted in Figure 1, could nevertheless be accommodated in this cell without spilling over into the "experimental" cell (or one of the others).

Other forms of inquiry fit into either of the remaining cells. For example, Piaget's clinical methods of studying children (i.e., posing certain standard situations and observing children's reactions to them) seem to fall into the lower right cell: high manipulation of antecedent conditions but no con-

straints imposed on outputs. Many simulation situations (and that old party standby–improvisations) also seem to be classifiable here. Observation schedules can be assigned to the upper left hand cell (e.g., the Flander's Interaction Analysis format which permits recording of teacher-pupil classroom interactions in terms of a systematic format without, however, prescribing the nature of the activities to be observed).

The definition of naturalistic inquiry in terms of the dimensions indicated in Figure 2 will be the standard for this monograph. All subsequent remarks and interpretations should be understood in these terms.

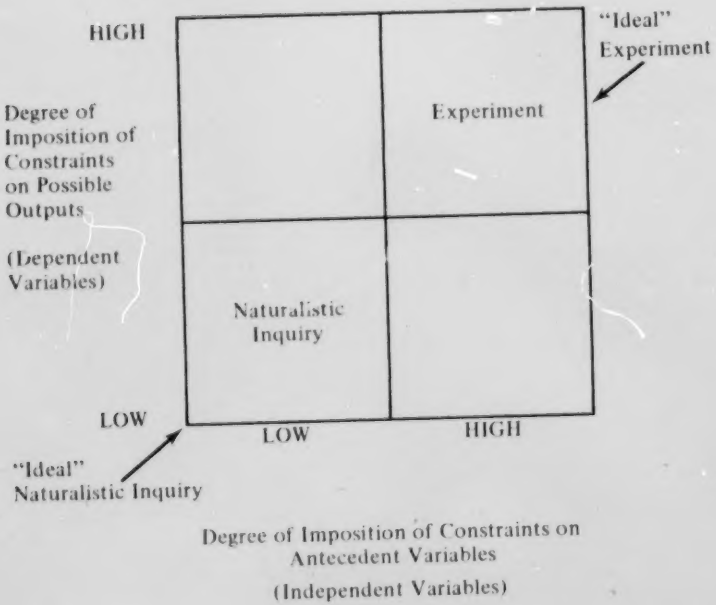

Figure 2: Representation of Inquiry Space or Domain of Inquiry

Based on Willems & Raush, 1969, p. 47

HOW NATURALISTIC INQUIRY DIFFERS FROM MORE CONVENTIONAL INQUIRY

The definition of the preceding section suggest that N/I differs from more conventional inquiry as a function of the constraints placed on antecedent conditions and on outputs by the investigator. But that distinction provides no basis for speculating about why an investigator would prefer N/I under some conditions.[6] To understand that, one needs to turn to other considerations.

The discussion which follows will present fourteen ways in which the naturalistic inquirer and the conventional inquirer tend to see the world from quite different perspectives. To begin with, however, it may be useful to note that there are forms of inquiry that are not well encompassed by the rubrics posed in Figure 2. For example, historical method or philosophical method cannot be fitted into the depicted inquiry domain without doing these methods some injustice. Inquiries in these areas can *sometimes* be understood within the framework of Figure 2. For example, it is possible to pursue the study of history either experimentally (by looking for so-called naturally occurring combinations of constraints which the historian would impose if he were able) or naturalistically (by analysis of documentation and records including personal papers). But most instances of these inquiry forms do not fit readily into the Figure 2 dimensions. In order to avoid such peripheral issues, however, let it be stipulated that the comments to follow will be directed toward primarily empirical inquiries that require interaction with human subjects. Most social, behavioral, and educational inquiry can be subsumed under such a definition.

Naturalistic inquiry can be contrasted with *conventional inquiry* (which ordinarily is experimental inquiry, although such inquiry forms as survey research are certainly included) in at least these ways:

1. ***Philosophical base.*** Bogdan and Taylor (1975) differentiate the two relevant philosophical perspectives nicely:

> Two major theoretical perspectives have dominated the social science scene. One, *positivism*, traces its origins to the great social theorists of the nineteenth and early twentieth centuries and especially to Auguste Comte and Emile Durkheim. The positivist seeks the *facts* or *causes* of social phenomena with little regard for the subjective states of individuals. Durkheim advised the social scientist to consider "social facts," or social phenomena, as "things" that exercise an external and coercive force on human behavior.
>
> The second theoretical perspective, which, following the lead of Irwin Deutscher, we will describe as *phenomenological*, stems most promi-

[6]Related analyses may be found in Patton (1975), Parlett and Hamilton (1976), and Shrock (1977).

nently from Max Weber. The phonomenologist is concerned with *understanding* human behavior from the actor's own frame of reference . . . . The phenomenologist examines how the world is experienced. For him or her the important reality is what people imagine it to be.

Since the positivists and the phenomenologists approach different problems and seek different answers, their research will typically demand different methodologies. (p. 2)

In a similar vein Cronbach (1975) writes,

> The positivistic strategy of fixing conditions in order to reach strong generalizations . . . fits with the concept that processes are steady and can be fragmented into nearly independent systems. Psychologists toward the physiological end of our investigative range probably can live with that as their principal strategy. Those of us toward the social end of the range cannot. (p.123)

Instead, Cronbach suggests that psychologists employ what he calls correlational research:

> Correlational research is distinguished from manipulative research in that it accepts the natural range of variables, instead of shaping conditions to represent a hypothesis. (p.124)

Thus, a first, and probably crucial, difference between naturalistic and conventional inquirers is their philosophical base. The naturalistic investigator is a phenomenologist while the conventional inquirer is a logical positivist. The phenomenologist is concerned with describing and understanding social phenomena; the positivist with "scientific" facts and their relationship to one another.

2. *Inquiry paradigm.* A second difference between naturalistic and conventional inquirers can be found in the paradigm that guides inquiry. The conventional investigator, with his positivist leanings, tends to see the world as composed of variables. Certain of these variables, called independent variables, can be manipulated to determine their effects on other, dependent variables. The purpose of the laboratory is to rid the field of certain variables not of interest but having possible effects which would produce erroneous results. In the physical sciences it is almost always possible to effect physical control over such confounding variables, but control is far more difficult to manage in dealing with human behavior. Here the investigator must resort to a randomization strategy which, while not eliminating the confounding variables, at least permits him to estimate their size through statistical processes. Thus, the basic paradigm is (a) identify the independent and dependent variables of interest (sometimes called the information variables), (b) eliminate possible confounding variables through laboratory controls to whatever extent resources and circumstances permit, (c) randomize selection and assignment of subjects to treatments, thus making possible the estimation of

the effects of confounding variables (determination of the error term), and (d) compare the effect in the dependent variable with the error term to reach a judgment of its probable reality. The essential model for this process is experimental physics; the tactics are those of laboratory control when possible and statistical manipulation when not.

The naturalistic investigator uses a quite different paradigm. He is concerned with description and understanding; thus, he begins as an anthropologist might begin learning about a strange culture, by immersing himself in the investigation with as open a mind as possible, and permitting impressions to emerge. As impressions are formed, he checks them out by various means, e.g., "triangulation," testing one source against another until he is satisfied that his interpretation is valid. Essentially the naturalistic inquirer's model is ethnography. In a more contemporary vein, the naturalist might claim his model to be investigative journalism, in which "truth" can be elicited from partial and even reluctant sources by processes of cross-checking, triangulation, and re-cycling until convergence is achieved.

3. *Purpose.* A third difference between naturalistic and conventional inquirers is purpose. The conventional inquirer deals with variables and their relations; and when he approaches an inquiry it is, typically, to test some proposition about a relationship called a hypothesis. His purpose is thus essentially that of verification, a testing of ideas in some empirically elaborated form.

The naturalistic inquirer on the other hand has as his purpose the *discovery* of phenomena whose empirical elaboration and testing would be worthwhile. However, that is not to suggest that he will not at times operate in a verification mode. When he has discovered interesting phenomena he may also wish to test their relationships. But even then he will continue to operate naturalistically, looking for instances in which the relationship can be observed rather than arranging for it to happen under controlled conditions. As will be seen, the issue of generalizability is very much at stake; naturally occurring relationships are much more likely to be observed again than those effected under laboratory conditions.

4. *Stance.* A fourth difference that may be noted is in the stance of naturalistic and conventional inquirers. The latter take essentially a *reductionist* stance; indeed, much of the reduction is accomplished prior to the actual investigation through the processes of imposing constraints on antecedent conditions and/or on outputs. Thus the conventional inquirer enters the field with preformulated questions or hypotheses, and he seeks only that information which will answer those questions or test those hypotheses.

On the other hand the naturalistic inquirer takes essentially an *expansionist* stance. He seeks a holistic view that will permit him to describe and

understand phenomena as wholes, or at least in ways that reflect their complexity. He enters the field and builds outward from wherever he happens to find himself. Each step in his inquiry depends on the sum of his insights from previous steps.

Thus, the conventional inquirer takes a structured, focussed, singular stance, while the naturalistic inquirer tends to be open-minded, exploratory, and complex in his position.

5. *Framework/design.* A fifth difference concerns the framework or design employed by each of the two types of investigators. Given what has been said about conventional inquiry so far, it is plain that in this mode designs must be constructed *preordinately,*[7] i.e., before the fact; or, as the root of the term suggests, they must be preordained. Moreover, once a design has been put into action, it is essential that it not be altered in any way, since such an alteration could only further confound the variables and their relationships and make an integrated interpretation of findings impossible. Thus, *preordinate, fixed* designs are one of the hallmarks of conventional inquiry.

For the naturalistic inquirer the design can be given in advance only incompletely; to specify it in detail would place constraints on either antecedent conditions or outputs or both, thereby altering the nature of the inquiry from naturalistic to conventional modes. In the N/I situation, the design emerges as the investigation proceeds; moreover, it is in constant flux as new information is gained and new insights are formed. Thus, *emergent, variable* designs are among the hallmarks of naturalistic inquiry.

6. *Style.* The sixth difference is the style with which the investigator "comes upon" the elements to be studied. The basic style of the conventional inquirer is *intervention.* He manipulates the situation so as to arrange those conditions of the dependent and independent variables that he wishes to observe. (Indeed, proper manipulation is considered essential to the establishment of the internal validity of the inquiry.) Ideally, as has been noted, the conventional inquirer mounts his inquiry in a laboratory setting.

The basic style of the naturalistic inquirer is *selection.* He does not manage the inquiry situation but uses it; he is less a stage manager than a member of the audience. He watches the entire play and then selects from it those aspects which he considers critical for his purposes. It may take a long time for the right combination of factors to occur, but that time is the price one must pay to be able to study phenomena under truly natural conditions.

Operationally, it is likely that the distinction between intervention and selection best differentiates conventional from naturalistic investigation. Any

[7] I am indebted to Robert Stake for this term, which he has coined to contrast with the responsive model of evaluation currently associated with his name.

intervention, any manipulation, immediately characterizes an inquiry as more experimental than naturalistic.

7. *Reality manifold.* A seventh distinction that may be drawn concerns the inquirer's view of reality. The conventional inquirer deals with a single, objective reality which he believes to exist and which his methods can uncover. Truth is truth, and when it is known, it is logically compelling. There is only one reality, and the experimentalist is its prophet.

The naturalistic inquirer is likely to take a much more flexible view of the nature of reality. There are at least three reasons for doing so. First, much of the reality with which the naturalistic inquirer must deal exists only in the minds of individual people and depends heavily on their separate perceptions. It is not surprising that their views of reality should differ. But who is to say who is right and who is wrong? The conventional criterion in this case is intersubjective agreement, but as shall be seen later in this monograph, this criterion is suspect at best.

Second, the naturalistic inquirer recognizes that the reality manifold is constantly changing in terms of time, people, episodes, settings, and circumstances. One should expect "reality" to be different at different times, recognizing the difference will depend on the situation and not necessarily or merely on a lack of reliability in methodology.

Finally, naturalists understand that reality, like the onion, has many layers. Eisner (in press), in discussing a form of evaluation which he terms educational connoisseurship, points out that

> The need for unanimity among critics is not characteristic of criticism since it is recognized that complex phenomena–works of literature, painting, film, and the like–have several layers of meaning and that the greatest works seem inexhaustible in the meanings one can secure from them. What is sought is not the creation of one final definitive criticism of a work, but rather the goal of having our perception and understanding expanded by the criticism we read. Classrooms and schools are at least as multilayered as works of art and we should seek, therefore, not a single definitive criticism, but rather criticism that is useful. Indeed, we should anticipate that critics with different educational orientations and interests will find in situations as phenomenologically dense as classrooms different things to describe, interpret, and evaluate. The cultivation of such productive diversity is a virtue, not a vice. Like education itself we do not seek to create an army marching in step to the same tune, but individuals who follow their own drummer as long as the beat is interesting.

Thus, the conventional inquirer may be said to seek a singular reality, while the naturalistic inquirer seeks a multiple reality.

8. *Value structure.* An eighth difference that may be noted concerns the value structure which each type of inquirer perceives. The situation here is

parallel to that just described for realities, and might in fact be considered part of the reality manifold. Since values are of such import in evaluation, however, and since value conflicts are so much a part of the contemporary social scene, it is probably useful to make a special point of the difference.

The conventional inquirer tends to view the world through a single value framework. Indeed, conventional inquirers have historically acted as if inquiry were, and could be, value free. Values could of course be taken as variables, and differences in value positions of different individuals and groups could be studied, but the inquirer himself had no values that were involved in the study.

The naivete of that position is now well known, even in the physical sciences. In educational evaluation, new models based on positions of advocacy rather than value freedom are emerging (see Chapter 3).

The naturalistic inquirer recognizes that his own values are very much part of his inquiry and that he needs to be as explicit about them as he can, both to avoid misleading persons who use his findings as well as deluding himself. But more importantly, he recognizes that there are multiple value positions that can be taken and that these are not only worthy of description but also form a dynamic element in the study itself. The value positions of different parties must be understood and honored if the study is to be authentic.

Thus, the conventional inquirer tends toward a single value position, while the naturalistic inquirer tends toward a pluralistic view.

The eight differences which have been discussed so far represent what one might call conceptual or theoretical differences. The six which follow have a somewhat more operational or practical orientation.

9. *Setting.* For various reasons already discussed it is clear that the conventional inquirer leans toward the laboratory setting for his investigations, while the naturalistic inquirer carries out his inquiry in a natural, i.e., non-contrived, environment.

10. *Context.* The conventional inquirer wishes to render the context as unrelated as he can to the inquiry, to screen out uncontrolled variation. To accomplish this end, he withdraws to the laboratory—the laboratory is, after all, the epitome of the context-free environment. The naturalistic inquirer finds the context entirely relevant—indeed, the phrase, "context-free naturalistic inquiry," is a logical contradiction. The naturalistic inquirer, far from screening out context, makes every effort to understand it so that he can assess its meaning for and impact on the elements being studied.

11. *Conditions.* The conventional inquirer seeks to control conditions; the naturalist opens his inquiry to uncontrolled conditions as much as possible. Indeed, he seems to *invite interference*[8] so that he can understand how events occur in the "real" world. As will be seen, the concept of invited interference is of great importance to the evaluator, who generally does not wish to know

[8]This phrase was coined by the late C. Ray Carpenter.

how the entity that is being evaluated works in the *best* of all possible worlds, but in the *worst*.

12. *Treatment.* The concept of treatment is extremely important in experimental science. Very often evaluators conceptualize the entity they are evaluating, e.g., a new school program, as a treatment because of the convenience this formulation affords in thinking about evaluation as though it were experimental research. Of course the treatment in any experiment must be stable and invariant, otherwise it is impossible to determine the effect uniquely associated with a given cause.

To the naturalist the concept of treatment is very foreign since it implies some kind of manipulation or intervention by the inquirer. If it should occur to him to consider some naturally occurring phenomenon as a "treatment," i.e., a likely cause for some observable effect, he would certainly not expect stability. The naturalistic inquirer realizes that continuous change is the essence of real situations.

13. *Scope.* The conventional inquirer must necessarily focus on a limited range of variables to be able to deal with them in the controlled, systematic way that characterizes his approach. One might suggest that the scope of his inquiry is molecular. Conversely, the naturalistic inquirer is more ready to consider *any* variable (should he think of inquiry in that way) that appears relevant. The scope of his inquiry is probably best characterized as molar.

14. *Methods.* Both conventional and naturalistic inquirers wish to be objective in their methodology, but the meaning which they ascribe to that term is quite different. The conventional inquirer strives for objectivity in the sense of *inter-subjective agreement*, i.e., agreement among two or more equally competent observers. The naturalistic inquirer, for a variety of reasons, not the least of which is the belief in a multiple reality, places little store in that form of objectivity and strives instead for *confirmability*, i.e., agreement among a variety of information sources. This issue of objectivity is of major importance and will be treated in some detail in a later chapter of this monograph. For the time being, however, it may simply be noted that this difference does exist among conventional and naturalistic inquirers.

The fourteen points of difference noted in the foregoing discussion are summarized in Table 1. It would be naive to believe that every conventional inquirer, under all circumstances, would cling to the beliefs ascribed to the view in this table, just as it would be absurd to suppose that the naturalistic inquirer never deviates from *his* most basic positions.

Nevertheless, the dimensions of the table, stated in this bald and polar form, well illustrate the fundamental differences in points of view between conventional and naturalistic inquirers. These differences determine the position that one takes in the inquiry field defined in Figure 2. Both conventional and naturalistic investigators place themselves in a particular part of that

Table 1: Some Basic Differences Between Conventional and Naturalistic Inquiry

Points of Comparison	Forms of Inquiry: Conventional Inquiry	Forms of Inquiry: Naturalistic Inquiry
Philosophical base	Logical positivism	Phenomenology
Inquiry paradigm	Experimental physics	Ethnography; investigative journalism
Purpose	Verification	Discovery
Stance	Reductionist	Expansionist
Framework/design	Preordinate/fixed	Emergent/variable
Style	Intervention	Selection
Reality manifold	Singular	Multiple
Value structure	Singular	Pluralistic
Setting	Laboratory	Nature
Context	Unrelated	Relevant
Conditions	Controlled	Invited interference
Treatment	Stable	Variable
Scope	Molecular	Molar
Methods	Objective–in sense of inter-subjective agreement	Objective–in sense of factual/ confirmable

inquiry field *because* of the views they hold on the Table 1 dimensions; they do not hold those views because of the place in which they find themselves in the inquiry domain.

HOW NATURALISTIC INQUIRY IS SIMILAR TO OTHER METHODOLOGIES

We have already noted the strong similarity of naturalistic inquiry to ethnography, which is the precursor and paradigm for N/I. But there are other, newer methodologies, which illustrate the fact that naturalistic approaches are alive and well in a number of contemporary settings:

1. *Grounded theory*. The term "grounded theory" was given currency in sociological inquiry by the publication in 1967 of the book, *The Discovery of Grounded Theory*, by Glaser and Strauss. Subtitled, "Strategies for Qualitative Research," the book defines the meaning of, and proposes strategies for, inquiry that is theory-*generating* as opposed to theory-*testing*. Grounded

theory may be contrasted to any hypothetico-deductive theory which is based on *ungrounded* assumptions, i.e., assumptions that are typically not data-based and may be generated by a variety of non-empirical means. Glaser and Strauss (1967) argue that:

> In contrasting grounded theory with logico-deductive theory and discussing and assessing their relative merits in ability to fit and work (explain and be relevant), we have taken the position that the adequacy of a theory for sociology today cannot be divorced from the process by which it is generated. Thus one canon for judging the usefulness of a theory is how it was generated—and we suggest that it is likely to be a better theory to the degree that it has been inductively developed from social research.
>
> Generating a theory from data means that most hypotheses and concepts not only come from the data, but are systematically worked out in relation to the data during the course of the research. *Generating a theory involves a process of research.* By contrast, the *source* of certain ideas, or even "models," can come from sources other than the data. . . . But the generation of theory from such other sources must then be brought into relation to the data, or there is great danger that theory and empirical world will mismatch. (pp. 5-6)

It is apparent that the grounded theorist bears a much closer relation to the naturalistic inquirer than to the conventional investigator. While most naturalistic inquirers are probably not as firmly wedded to the idea of theory generation as are Glaser and Strauss, the basic methodologies are similar. In particular, it is plain that the grounded theorist is interested in discovery rather than verification, a point noted as distinguishing the two forms of inquiry considered in the preceding section.

2. *Ecological psychology.* Roger Barker is generally called the father of ecological psychology because of his pioneering work in this field. While a number of his ecologically-oriented papers preceded the publication of the paper, "Explorations in Ecological Psychology," in 1965, it is that article which brought widespread attention to his ideas. A review of some of these ideas is enlightening.

Empirical phenomena, Barker points out, occur without the benefit of intervention by an investigator. However, data relating to those phenomena are always the product of an interaction between the phenomena and the investigator. Two kinds of interaction are possible. In the first case, called "Type *T*," the relationship between phenomena and data is transitive, i.e., the data *pass through* the investigator who simply codes them so that they may be analyzed later at the investigator's convenience. In the second case, called "Type *O*," the relationship is looping; the investigator becomes an integral part of the data through intervention. Essentially, Barker suggests that *T* systems are natural and uncontrived, while *O* systems are controlled and manipulated.

Barker (1965) provides a noteworthy example of the difference between these two modes:

> Some years ago, when I was a student of Kurt Lewin, he and Tamara Dembo and I carried out some experiments on frustration. The findings of these experiments have been verified by others and they have become a part of the literature of scientific psychology. The experiments provided basic information about the consequences for children of frustrations, as defined in the experiments, and about the processes that produce these consequences. Time passed. In due course *I* had a student, and he undertook to study frustration. So far, so good. My student, Clifford L. Fawl, did not replicate the earlier study; he did not *contrive* frustration for his subjects; he pioneered, and extended the investigation from children *in vitro*, so to speak, to children *in situ*. He searched our specimen records of children's everyday behavior for instances of this allegedly important phenomenon without psychologists as operators. Here are the words of his report:
>
> > The results . . . were surprising in two respects. First, even with a liberal interpretation of frustration fewer incidents were detected than we expected. . . . Second, . . . meaningful relationships could not be found between frustration . . . and consequent behavior such as . . . regression . . .' and other theoretically meaningful behavioral manifestations.
> >
> > In other words, frustration was rare in the children's days, and when it did occur it did not have the behavioral consequences observed in the laboratory. It appears that the earlier experiments simulated frustration very well as we defined it and prescribed it for our subjects (in accordance with our theories); but the experiments did not simulate frustrations as life prescribes it for our children (p. 5).

Equating Barker's Type *T* systems with naturalistic inquiry and his Type *O* systems with experimental inquiry would not be far off the mark. In the fifteen or so years since its initial emergence, ecological psychology has attracted many adherents; it may be the best established version of naturalistic inquiry in existence.

3. *Aexperimental inquiry in educational change.*[9] A conference on strategies for educational change was held in Washington in 1965, on which occasion a paper was delivered by the present author entitled, "Methodological Strategies for Educational Change." In this paper, a distinction was drawn between experimental and "aexperimental" inquiry, the latter characterized as essentially observational and non-interventionist. The two modes were contrasted on a number of characteristics, including the investigator's intent, the setting, the level of control, the scope of inquiry, the number of variables involved, the treatment, and the context—a presentation not dissimilar from, although considerably more truncated than, that of Table 1. After some

[9]An example of the application of "aexperimental" methodology as described here may be found in O'Keefe (1968).

discussion of the nature of the change process, which was intended to show that aexperimental approaches seem more appropriate than experimental approaches in studying change, a number of strategies were suggested for field studies of change. These strategies included programmatic approaches which were conducted within an explicit logical framework (a conceptualization of the change process), continuous time analysis, constant replication and recycling, purposeful sampling, and heavy reliance on logical rather than statistical inference (although quasi-designs were suggested as useful in circumstances of "naturally-occurring" experiments).

The analog between aexperimental inquiry and naturalistic inquiry is apparent. While the community of change process researchers did not seize upon the suggestions made, it is clear that naturalistic strategies are highly appropriate to such investigations, and that the opportunity to apply them continues to exist.

4. *Educational anthropology/ethnography.* One of the roots of naturalistic inquiry is deeply buried in its precursor fields of anthropology-ethnography. That these two fields should currently be displaying a good deal of vigor in the educational R&D community should therefore come as no surprise. Instead this vigor may be taken as evidence that naturalistic approaches are beginning to be appreciated.

A recent overview of this emergent field by Wilson (1977) reveals a close association between the anthropological/ethnographic approach and naturalistic inquiry. In a summary section Wilson asserts:

> We have briefly explained, then, the two sets of hypotheses underlying the rationale for ethnographic research: (a) Human behavior is complexly influenced by the context in which it occurs. Any research plan which takes the actors out of the naturalistic setting may negate those forces and hence obscure its own understanding. (b) Human behavior often has more meaning than its observable "facts." A researcher seeking to understand behavior must find ways to learn the manifest and latent meanings for the participants, and must also understand the behavior from the objective, outside perspective. (p. 253)

Wilson goes on to assert that "these hypotheses taken together fundamentally challenge the way that much traditional research is conducted," causing debate that "can be beneficial if it leads researchers of all persuasions to question their basic assumptions about human behavior and ways of understanding it." (p.253)

Actual applications of modern ethnographic techniques in educational research are still rare. A 1974 article by Lutz and Ramsey advocating their use in education contains no references in its bibliography to such studies.[10] Since then a few studies have been published; for example, the entire May, 1977 number of *Anthropology and Education Quarterly* was devoted to

[10]Although some might have been mentioned.

"Exploring Qualitative/Quantitative Research Methodologies in Education." Among the more interesting articles in this special issue were those of Erickson, outlining some techniques for ethnographic study in schools; Fienberg, attempting to reintroduce some concepts of quantitative data collection and analysis into ethnographic approaches; and Herriott, reporting an actual ethnographic study done in the context of the Experimental Schools program. Contributions by Jackson (1968), Rist (1975), and Wolcott (1977) are also notable.

The practitioners of educational anthropology/ethnography probably cannot be understood as a homogeneous school. Investigators are taking different approaches, and in any event, the field is too young to permit the identification of definitive tendencies. Nevertheless, it is apparent that a vital field is emerging, one which on principle is closely allied to the tenets of naturalistic inquiry.

5. *Ecological experiments.* Urie Bronfe[illegible] (1976, 1977) has proposed the concept of the ecological experimen[illegible]uch might be viewed as an effort to bring classical experimentalism into the naturalistic camp. Bronfenbrenner (1977) begins his argument by suggesting:

> To corrupt a contemporary metaphor, we risk being caught between a rock and a *soft* place. The rock is *rigor*, and the soft place *relevance*. . . . The emphasis on rigor has led to experiments that are elegantly designed but often limited in scope. This limitation derives from the fact that many of these experiments involve situations that are unfamiliar, artificial, and short-lived and that call for unusual behaviors that are difficult to generalize to other settings. From this perspective, it can be said that much of contemporary developmental psychology is *the science of the strange behavior of children in strange situations with strange adults for the briefest possible periods of time.* (p. 513)

To avoid this problem, Bronfenbrenner asserts, the experiment should not be used as a verifactory tool but as a *heuristic strategy*:

> It is from this perspective that the primary purpose of the ecological experiment becomes not hypothesis testing but *discovery*—the identification of those systems-properties and processes that affect, and are affected by, the behavior and development of the learner. (Bronfenbrenner, 1976, p. 6)

But the ecological experiment should not be thought of as equivalent to its classical, laboratory prototype:

> In the classical psychological experiment, antecedent and consequent conditions are couched in terms of variables that are conceived as linear, additive, and distinct from each other. In an ecological model, status and change are viewed not solely in terms of different levels of one or more separate variables, but also, and primarily, as differences in systems or system states. (Bronfenbrenner, 1976, p. 6)

Ecological experiements have many identifying characteristics, from the *setting* (a microsystem in which factors of place, activity, time, and role figure) through *meso-*, *exo-*, and *macro-systems*, the latter including the most over-arching cultural institutions. As described by Bronfenbrenner, ecological experiments have many similarities to naturalistic inquiry, including, for example:

- Ecological validity—real-life rather than laboratory settings.
- Contextual validity—time, place, activity, and role common to the subjects' subculture.
- Phenomenological validity—provision for determining and taking account of each participant's definition of the situation.
- Reciprocity—taking account of interactive effects not only of inquirer on participant but also of participant on inquirer.
- Indirect and interactive systems effects—taking account of social interactions and second- or higher-order effects produced by one sub-system or system (e.g., the schoolroom) on another (e.g., the home).
- Temporal and spatial arrangements—taking account of the physical and temporal context.

Thus the use of the term "experiment" by Bronfenbrenner is somewhat misleading, both in that it differs in *purpose* from the classic conception (discovery rather than verification) and that it differs in *implementation* (not highly focussed and carefully controlled but open to many influences and taking account of many factors). It seems likely that Bronfenbrenner, rather than mounting a frontal attack on his colleagues' values as bound up in the concept of the experiment, is advocating a side-stepping maneuver that retains the term "experiment" but assigns it a different meaning. While such a move may have political utility, it does little to distinguish the fundamentally different characteristics of the two approaches. The objective observer is likely to find a great many more parallels between ecological experiments and naturalistic inquiry than between ecological and classical experiments.

THE NEED FOR NATURALISTIC INQUIRY

In previous sections we have seen that naturalistic inquiry may be defined as occupying a portion of a domain or space in common with other inquiry modes (Figure 2), and we have contrasted N/I with conventional inquiry at the level of assumptions or overriding characteristics. But the mere fact that N/I can be defined into existence does not automatically mean that it is pre-eminent or even generally useful. Of course, no one would wish to argue that N/I ought to replace other forms of inquiry for all purposes; but what arguments might be advanced to suggest that N/I is a preferred mode in certain instances? This section will deal with several such arguments.

1. *To enlarge the arsenal of investigative strategies for dealing with emergent questions of interest.* There seem to be at least two reasons why such an enlargement is desirable:

a. Paradigms currently in wide use, largely borrowed from other disciplines, have had a rather unremarkable history in dealing with important social questions. Cronbach (1975) for example, has observed:

> Some 30 years ago, research in psychology became dedicated to the quest for nomothetic theory. . . . Model building and hypothesis testing became the ruling ideal, and research problems were increasingly chosen to fit that mode. Taking stock today, I think most of us judge theoretical progress to have been disappointing. Many are uneasy with the intellectual style of psychological research. (p. 116)

b. Currently used paradigms tend to ignore a. as of information of major concern and relevance in dealing with important questions. Consider the following statement by Barker:

> Scientific psychology knows nothing, and can know nothing, about the real-life settings in which people live in ghettos and suburbs, in large and small schools, in regions of poverty and affluence. One might think that in the course of its necessary concern with stimuli, psychology would have become informed about the human environment. But this is not the case. Psychology has necessarily attended to those elements of the environment that are useful in probing its focal phenomena, namely, the behavior-relevant circuitry within psychology's black box. Psychology knows much about the physical properties and dimensions of the environmental probes it uses—of distal objects of perception, for example, and of energy changes at receptor surfaces. But the problem is that, in the course of its investigations, it has excised these environmental elements from the contexts in which they normally occur: mealtimes, offices, airplanes, arithmetic classes, streets and sidewalks.
>
> In view of psychology's concern with such dismantled fragments of its environment, it is not surprising that general conceptions of the environment occupy a minor place in the science, and that these conceptions provide a distorted view of intact settings in which behavior occurs. (Willems & Raush, 1969, p. 31)

2. *To provide an acceptable basis for studying process.* Psychology's "black box" has a unique property—one cannot see what goes on inside it. The black box concept is borrowed from physics, in which processes often cannot be viewed or even grasped. It was proposed that physics be studied as though it took place in a tightly sealed and inaccessible "black box," which had various knobs, levers, wheels, pointers, and dials sticking out of it. The object of the game was to determine how any given manipulation of a knob, lever, or wheel might influence, say, a dial or pointer reading; and then to guess what must be going on inside the box from the external evidence of these manipulations. (One may note the parallel of the black box to the

experimental model which requires manipulation of antecedent conditions and constrained output readings.) But it should be observed that this strategy was espoused precisely because the box and its mysterious internal processes were inaccessible to direct observation; had the physicist been able to open the box he surely would have done so at once.

As a general premise, it is probably safe to assert that the best way to study process is to observe it directly, rather than to infer its nature from the known input and the observed output. When process is the issue, naturalistic inquiry seems to offer a more useful means for its study than does the experimental mode.

3. *To provide an alternative where it is impossible to meet the technical assumptions of the experimental approach in the real world.* The power of the experimental approach derives precisely from the control which it is able to exercise over all relevant variables, either physically or statistically. In the laboratory it is the method *par excellence*. But when experiments are attempted in real world settings, the method suffers because so many of its assumptions, such as random selection and assignment, treatment additivity, or invariant context, cannot be met. Gene Glass has observed[11] that field experiments (or planned variations as they are sometimes called) provide little opportunity to discover anything useful because they rest on three probably invalid assumptions: that methods are singular (consistent from site to site), that true differences do occur among methods being compared, and that if outcome measures are appropriate (mainly reliable), experiments will be definitive. But methods are rarely singular. (Consider for example, the enormous site-to-site treatment variations that occur in a Federal program such as Follow-Through, or the effects of method-situation interaction.) Whether or not there are true method differences is precisely the issue; such differences cannot be assumed. Finally, since the definitiveness of an experiment rests in turn on treatment invariance and the existence of true differences, the third assumption also tends to be unmet. In another context, Alice Rivlin (1973) after massive experience in the evaluation of Federal programs, indicates her doubt that gross experimental comparisons can ever produce findings definitive for schooling.

The reader may protest that the inability to meet assumptions is not the fault of the method but of the sloppy way in which it is applied in the field. If the implication of such an assertion is that these problems could be eliminated by more rigorous application, then the asserter is simply ignorant of the facts of life in field studies. The array of human, political, and social functions, to say nothing of the basic intractability of conditions, will *always*

[11]Comments at a meeting of the Working Group on Educational Evaluation, Boulder, Colorado, May 17, 1977.

cause the experimenter to fall short of the ideal; and to the extent that he does, the findings are proportionately less credible.

4. *To better assess the implications of treatment-situation interaction.* Conventional inquiry tends to treat interaction effects, as for example, the familiar aptitude-treatment interaction (ATI), as "higher-order" and therefore relatively insignificant.[12] This tendency is probably based on the analogous tendency among physical scientists to make "simplifying assumptions" when faced with otherwise unsolvable problems, or problems which, although solvable, would have a solution form too complex for ready use in practice. So for example, the "three-body" problem is essentially unsolvable unless one assumes that one of the bodies is very much larger than the other two (the case, for example, of the motions of airplane, catapulted pilot, and earth is solvable). Or, description of the motion of a pendulum requires a complex equation unless one is willing to assume that the angle of displacement of the pendulum is small—less than six degrees, in fact, because the nasty term in the equation involves $\sin \theta / \tan \theta$, which is essentially 1 for small angles. In these examples such simplifying assumptions are warranted and useful, but in many behavioral cases they are not. Cronbach (1975) indicates that:

> Social scientists generally, and psychologists in particular, have modeled their work on physical science, aspiring to amass empirical generalizations, to restructure them into more general laws, and to weld scattered laws into coherent theory. That lofty aspiration is far from realization. A nomothetic theory would ideally tell us the necessary and sufficient conditions for a particular result. Supplied the situation parameters A, B, and C, a theory would forecast outcome Y with a modest margin of error. But parameters D, E, F, and so on, also influence results, and hence a prediction from A, B, and C alone cannot be strong when D, E, and F vary freely. Theorists are reminded from time to time that the person who states a principle must also state the boundary conditions that limit its application. The psychologist can describe the conditions under which his generalizations have held, or the domain of which they provide an actuarial summary. He cannot often state the boundaries defining how far they will hold. (p. 116)

Higher-order interactions are important in social inquiry, and probably particularly in evaluative inquiry. Naturalistic inquiry seems to have some edge because of its emphasis on detailed descriptions of such interactions; it may thereby facilitate some understanding of their dynamics.

5. *To redress the balance between reconstructed logic and logic-in-use.* Chris Argyris (1975) draws a distinction between "espoused theories of action . . . that people report as theories that inform their actions and theories-in-use . . . inferred from how people actually behave" (p.469). This

[12]This allegation is of course untrue for the work of Cronbach, Snow, and other distinguished investigators in this field.

distinction is similar to that made by Abraham Kaplan (1964) between reconstructed logic and logic-in-use:

> Now the word "logic" is one of those, like "physiology" and "history," which is used both for a certain discipline and for its subject matter. We all have physiologies and histories, and some of us also think and write about these things. Similarly, scientists and philosophers use a logic—they have a cognitive style which is more or less logical—and some of them also formulate it explicitly. I call the former *logic-in-use* and the latter the *reconstructed logic*. We can no more take them to be identical or even assume an exact correspondence between them, than we can in the case of the decline of Rome and Gibbon's account of it, a patient's fever and his physician's explanation of it. (p. 8)

Thus, there is a difference between what people actually do as inquirers and what they report about what they do. The so-called scientific method is a reconstructed logic; it purports to describe what the ideal scientist does in pursuing an inquiry. That reconstruction probably has little connection with what a scientist actually does. His problem may manifest itself simply as a gnawing in his stomach and present itself in aspects quite different from those the scientist would choose could he follow his ideal method.

Unfortunately reconstructed logic often takes on the qualities of a myth or a religious belief. People come to believe that, in the pursuit of an inquiry, they actually behave as the model suggests they should (or at least they so pretend, perhaps in the interest of getting published). Worse, they come to believe that to act in any *other* way is unscientific and perhaps even unethical. It surely cannot lead to "truth."

The reconstructed logic of experimentalism has held sway in behavioral science inquiry for most of this century. This argument is not a plea to eliminate it; experimentalism has great power under the appropriate circumstances. It is a plea to recognize other methodologies as sometimes appropriate, to redress the imbalance that has occurred with such devastating results.

6. *To avoid implicit shaping of possible outcomes.* The outcomes of an inquiry are shaped by the interaction of two elements: what is "there" to be studied, and what the investigator introduces in the process of studying it. Conventional inquiry modes involve some degree of *a priori* manipulation. Regrettably, this manipulation, performed for the laudable purposes of control, may artificially determine the scope and sometimes even the specific findings of the inquiry.

Brunswik (1955) makes the case as follows:

> One of the most time-honored traditions in experimentally testing generalizability demands that one, or perhaps a few, conditions be varied in a planful manner decided upon by the experimenter while all others are held constant. The purpose is to assure isolation of the so-called independent variables. For their arbitrary orderliness and confinement such designs may be called "systematic."

> This intervention constitutes artificially induced perfect confounding, and may be labeled "tied-variables" design or, in short, tied design. (p. 194)

Commenting on this concept, Willems says:

> In Brunswik's terms, artificial tying occurs when variables are allowed or made to vary together in ways that persons never confront in their everyday lives, whereas artificial untying occurs when natural covariations of variables are eliminated through experimental control. (Willems & Raush, 1969, p. 55)

These statements suggest that conventional inquiry may actually be self defeating by artificially linking certain variables, making it inevitable that these variables will be linked in any findings. Similarly, the design may preclude the discovery of links which ought to be discovered. Tying or untying of variables cannot occur within naturalistic inquiry. When there is a risk of such tying or untying with important variables, naturalistic inquiry should be preferred.

7. *To optimize generalizability*. Conventional inquiry requires external manipulation on the part of the experimenter. Such manipulations may have seriously detrimental effects on the generalizability of the findings to real settings. Willems suggests:

> When an attempt is made to generalize from a manipulational study to an everyday phenomenon or a phenomenon that one has not produced, it is necessary to assume that the manipulated circumstances comprise a *model* of the everyday phenomenon, or that they *represent* it. The possible array of factors that limit the fit of the model or disrupt representativeness is . . . extensive. (Willems & Raush, 1969, p. 63)

In this context, it is useful to recall the comments of Glass (p.25) indicating the virtual impossibility of meeting the assumptions of the paradigm and its associated statistical techniques.

However, Cronbach (1975, pp.122-23) offers a much more compelling argument against generalizability in conventional inquiry. He suggests that generalizations, even those made under the "best" circumstances, are not terribly secure. After noting that recent studies on negative income tax lead to observations only about recent history and not to enduring conclusions, he gives a series of impressive examples of how what were seemingly sound generalizations can lose validity:

- The failure of DDT to control mosquitos as genetic transformations made them resistant to the pesticide.
- The shifting of stars in their courses so as to render star maps obsolete.
- The shifting in the value of the gravitational constant, so that while s will continue to equal gt^2, the actual distances covered by falling bodies will differ.

- The suggestion by Ghiselli that the superiority of distributed over massed practice may not remain valid from one generation to another.
- Atkinson's proposition that when a substantial relation is found between personality variables, that relation describes only the "modal personality" of a particular society at a particular time in history.
- Changes in the construct validity of the *California F Scale*.
- Bronfenbrenner's conclusion that class differences in parenting observed in the 1950s were just the reverse of those observed in 1930.

From this set of examples Cronbach (1975) concludes:

> Generalizations decay. At one time a conclusion describes the existing situation well, at a later time it accounts for rather little variance, and ultimately it is valid only as history. The half-life of an empirical proposition may be great or small. The more open a system, the shorter the half-life of relations within it are likely to be.
>
> . . . Propositions describing atoms and electrons have a long half-life, and the physical theorist can regard the processes in his world as steady. Rarely is a social or behavioral phenomenon isolated enough to have this steady-process property. Hence the explanation we live by will perhaps always remain partial, and distant from real events . . . and rather short lived. The atheoretical regularities of the actuary are even more time bound. An actuarial table describing human affairs changes from science into history before it can be set in type. (pp. 122-23)

What does Cronbach suggest as a counter-strategy?

> Instead of making generalization the ruling consideration in our research, I suggest that we reverse our priorities. An observer collecting data in the particular situation is in a position to appraise a practice or proposition in that setting, observing effects in context. In trying to describe and account for what happened, he will give attention to whatever variables were controlled, but he will give equally careful attention to uncontrolled conditions, to personal characteristics, and to events that occurred during treatment and measurement. As he goes from situation to situation, his first task is to describe and interpret the effect anew in each locale, perhaps taking into account factors unique to that locale or series of events. . . . As results accumulate, a person who seeks understanding will do his best to trace how the uncontrolled factors could have caused local departures from the modal effect. That is, generalization comes late, and the exception is taken as seriously as the rule.
>
> . . . When we give proper weight to local conditions, any generalization is a working hypothesis, not a conclusion. (pp. 124-25)

This stunning argument suggests that generalizability may be a holy grail, to be pursued but never quite attained. The best pursuit comes not from carefully controlled experiments presumed to yield permanently definitive results, but from very idiographic activity with the full range of local factors taken into account.

In all, we may agree with Willems and Raush's (1969) prescription that

> If generalization about a phenomenon is one of an investigator's aims, then his program of research should include naturalistic strategies. The ethological movement has this prescription as one of its basic canons. (p. 276)

8. *To meet, optimally, certain practical criteria for the derivations made from an inquiry.* Glaser and Strauss (1967, pp.3-12), in their discussion of grounded theory, discuss three criteria of this kind in terms of *fit, work,* and *communicate.* They use the term *fit* to mean that categories, theories, or interpretations derived from a study "must be readily (not forcibly) applicable to and indicated by the data." The term *work* indicates that such derivatives must be "meaningfully relevant to and able to explain the behavior under study." The term *communicate* implies that the derivatives should be sufficiently understandable to be used.

Naturalistic inquiry has a higher probability of success on these criteria than does conventional inquiry. It must be clear that derivatives (interpretations, categorizations, theories) will better fit the total data when resulting from naturalistic studies than will *a priori* stipulations which are tested by limited data segments. Relevance and the search for explanation are typical of N/I; hence, derivations have a high probability of working. Communication should also be well served—derivations from naturalistic studies are likely to take the form and reflect the context which the persons involved in the study know best—including decision-makers in an evaluative study.

Chapter III:
THE UTILITY OF NATURALISTIC INQUIRY FOR EDUCATIONAL EVALUATION

It is evident that naturalistic inquiry is a mode that can be applied in a wide variety of situations, including the full gamut of research that one might wish to pursue in educational or social/behavioral fields.[13] The remainder of this monograph, however, will focus on educational *evaluation* for a number of reasons:[14]

THEORETICAL MOVEMENT AWAY FROM CONVENTIONAL PARADIGMATIC MODELS

There is a strong interest among leading evaluation theorists in moving away from the more conventional paradigmatic models that have guided educational evaluation. This fact is perhaps nowhere better illustrated than in the recent in-depth review of the Guttentag-Struening *Handbook of Evaluation Research* done by the Stanford Evaluation Consortium (1976) under the editorship of Lee Ross and Lee J. Cronbach.[15] The Consortium suggests that there are two views of evaluation: the "mainstream" view, which currently dominates the field, and which, in terms of this paper, may be interpreted as based heavily on the conventional paradigm of inquiry, and an "alternative and extended view," which they characterize as follows:

> Reading through the *Handbook*, some limitations in the conventional or mainstream view of evaluation become evident. Several authors, like other contributors to the evaluation literature, speak to the need for one or another revision of the model--but no synoptic view of a better model emerges from the *Handbook*. We would emphasize the general features of such a model: (1) Evaluation can constructively enter the picture earlier and can be seen as a continuing part of management rather than as a short-term consulting contract, (2) The evaluator, instead of running alongside the train making notes through the windows, can board the train and influence the engineer, the conductor, and the passengers. (3) The evaluator need not limit his concerns to objectives stated in advance; instead, he can also function as a naturalistic observer whose inquiries

[13]Unfortunately, most of the questions currently being asked in such research are cast in forms that appear to require conventional inquiry modes to answer them.

[14]This focus is not as constraining as one might at first believe. While operationalized naturalistic inquiry will appear different in different fields, there are in fact many parallels and analogs which will occur to the reader.

[15]An interesting response to this review appears in Stake (1977).

> grow out of his observations. (4) The evaluator should not concentrate on outcomes; ultimately, it may prove more profitable to study just what was delivered and how people interacted during the treatment process. (5) The evaluator should recognize (and act upon the recognition) that systems are rarely influenced by reports received in the mail.
>
> Evaluation thus becomes a component of the evolving program itself, rather than disinterested monitoring undertaken to provide ammunition to the warring factions in a political struggle. Formal reports to outsiders are reduced in significance, and research findings become not conclusions but updates of the system's picture of itself. (p. 18)

While there are certainly some differences between the alternative and extended view identified by the Consortium and the naturalistic mode, there can be little doubt that N/I fits their concept better than does the experimentalism decried in the Ross-Cronbach review.

PRACTITIONER MOVEMENT AWAY FROM CONVENTIONAL EVALUATION METHODOLOGIES

There is a strong movement away from conventional evaluation methodologies toward more naturalistic approaches on the part of leading evaluation practitioners. Some examples of current work under way include:[16]

1. *The "School Without Schools" Project of Daniel Stufflebeam and James Sanders.* During the natural gas crisis occasioned by the severe winter of January-February, 1977, the Columbus, Ohio public schools were closed under a general order from the Governor of the State of Ohio in order to conserve fuel. The Columbus School District nevertheless resolved to provide some kind of educational experience for its youngsters, operating, in effect, school without schools. Daniel Stufflebeam and James Sanders of the Evaluation Center, Western Michigan University, were supported by the National Science Foundation to gather data on this experience.

Because of the transient nature of this situation it was impossible to pursue any detailed, systematic, preordinate evaluation strategy. Instead, extensive use was made of observations, interviews, tape recordings, video broadcasts, and other unusual data sources, in an effort to document the activities that occurred and to estimate their effects. The methodology was essentially naturalistic.

When the schools were reopened and the two evaluators had had an opportunity to reflect on their experience, they were sufficiently impressed with the outcomes of what surely would be termed non-rigorous, indeed, impressionistic inquiry, that they proposed a new entity to be called "Crisis Research in Education." The methodology for such research, which remains

[16] At this writing all these projects are still unpublished. The impressions reported here were gleaned from private communication with the evaluators cited.

to be explicated, would be utilized in other crisis situations of interest, e.g., ghetto riots, failures of bond issues which cause the closing of schools or severe curtailment of programs, busing upheavals, and other situations.

2. *The NSF-sponsored Case Studies of Science Education of Robert Stake and his associates.* Robert Stake, of the Center for Instructional Research and Curriculum Evaluation (CIRCE) of the University of Illinois, has undertaken to carry out a series of ten case studies on behalf of the National Science Foundation. He intends this effort to augment a national needs assessment in science which will guide NSF's policy decisions in the near future. Stake characterizes the methodology as a combination of naturalistic inquiry and survey approach, with the latter being used primarily to validate the case study findings over a broader base of communities. The inquiry is grounded largely on "non-interventionist observation" leading to "naturalistic generalizations." These generalizations form the basis for the survey instrument. Results are dependent on extended observations made by individual observers at each site (4-15 weeks) with each observer following whatever methodology and schedule that seem appropriate in view of the local situation.

3. *The OE/BEH-sponsored judicial evaluations of the Individual Educational Program (IEP) requirements of PL 94-142 conducted by Robert Wolf and Barbara Tymitz under contract with Nero & Associates, Inc.* In this study, the concerns and issues arising in the minds of relevant audiences with respect to the feasibility and effects of the IEP provision of PL 94-142 are being studied in four states using the judicial model of evaluation proposed by Robert Wolf.[17] Extensive interviews following naturalistic formats result in the identification of a series of issues which are then "tried" before a "jury" using the procedures developed in the law for administrative hearings. A distinguishing feature of this approach is its reliance on human testimony, with "circumstantial evidence" such as test scores being relegated to a secondary role (as they are in real courts of law).

The three cases described above are but examples of current tendency among leading evaluation practitioners to abandon more conventional inquiry forms in favor of naturalistic approaches. One must conclude that in this field the time is ripe for a more complete exposition of naturalistic methodology.

EMERGENT EVALUATION MODELS

A number of evaluation models have emerged which seem especially congenial to the use of N/I as a supporting tool or technique. Educational evaluation has undergone a major upheaval over the past ten to fifteen years. We are a long way from the time when the only approach to evaluation was the objectives-centered model explicated by Ralph Tyler (1950). After the

[17]This model will be described in more detail below.

surprise of Sputnik had stimulated large scale national curriculum projects such as PSSC Physics (NSF) and Project Social Studies (USOE), it was found that conventional evaluation approaches were simply inadequate to deal with the evaluation questions and issues posed by these vast projects.[18] New proposals began to emerge in earnest in 1967. At the time of this writing well over twenty models have made their way into the literature.[19] Five of these models seem especially congenial to the use of naturalistic inquiry as a major or supportive technique:

1. *The Responsive Model.*[20] Over the past decade Robert Stake has updated his well-known Countenance Model of evaluation (1967) into a form which he calls *responsive* evaluation, distinguishing it from the *preordinate* evaluation which he feels dominates current evaluation practice. There are, he suggests, many different ways of evaluating, and no one right way. But if one wishes to select an evaluation procedure that will emphasize those issues that are important for each program, one should select responsive evaluation. Evaluation is responsive "if it orients more directly to program activities than to program intents; responds to audience requirements for information; and if the different value perspectives are referred to in reporting the success and failure of the program." (Stake, 1975a, p.14)

The major purpose of evaluation should be to respond to audience requirements for information, particularly in terms of the value perspectives held by each audience. It is more important to improve audience understandings than to provide a "reductionist" summary. Indeed, the major criterion for validity of an evaluation is less the "truth-value" of its judgments than the degree to which audience understanding is increased.

The "advance organizer" for the evaluation, i.e., the particular focus for the information to be collected, is the *issues*, i.e., the problems or potential problems, perceived by the several audiences. Stake describes the process as follows:

> To do a responsive evaluation, the evaluator conceives of a plan of observations and negotiations. He arranges for various persons to observe the program, and with their help prepares brief narratives, portrayals, product displays, graphs, etc. He finds out what is of value to his audiences, and gathers expressions of worth from various individuals whose point of view differ. Of course, he checks the quality of his records: he gets program personnel to react to the accuracy of his portrayals; and audience members to react to the relevance of his

[18]See Cronbach (1963) for an excellent discussion of these questions and issues.

[19]Among the more prominent of these models are the CIPP Model of Stufflebeam and his associates (1971); the Countenance Model of Robert Stake (1967); the Discrepancy Model of Malcolm Provus (1971); and the Goal Free Model of Michael Scriven (in House, 1973).

[20]The major references for responsive evaluation are Stake (1975a) and (1975b).

> findings. He does much of this informally–iterating and keeping a record of action and reaction. He chooses media accessible to his audiences to increase the likelihood and fidelity of communication. He might prepare a final written report, he might not–depending on what he and his clients have agreed on. (1975a, p. 14)

It is the series of steps that must be taken to carry out a responsive evaluation that make it plain why there is a natural affinity between this approach to evaluation and naturalistic inquiry. These steps (not necessarily in this order and with frequent recycling) direct the evaluator to: talk with clients, program staff, and audiences; identify program scope; overview program activities; discover purposes and concerns; conceptualize issues and problems; identify data needs relevant to the issues; select observers and instruments (if any); observe designated antecedents, transactions, and outcomes; thematize–prepare portrayals and case studies; winnow, matching issues to audiences; format for audience use; assemble formal reports (if any). (Stake, 1975a p.20)

Several phrases indicate the naturalistic bent of Stake's responsive evaluation:

- *overview program activities* (enter the field with a minimum of prior manipulation.)
- *discover purposes and concerns* (rather than verify purposes and concerns that have been deduced from *a priori* formulations such as statements of objectives).
- *conceptualize issues and problems* (derive categories that fit the information that has been obtained.)
- select observers and instruments *if any* (clearly the precise measurement of variables is a secondary consideration.)
- *thematize* (portrayals and case studies are better vehicles for understanding the data than reports.)
- prepare formal reports *if any* (formality takes a back seat to effective communication.)

The evaluator's task is first to focus on issues and then to collect information relevant to them (even test scores if they should prove to be important). The evaluator continuously portrays his findings to his audiences, to be sure that there is effective communication and increased understanding. So far as possible the audiences are provided with *experiences*–real or vicarious–in order to have firsthand impressions of what things are really like. If reports are written, it is the detail of vicarious experience that substitutes for the reductionist summary usually found in such reports.

This approach seems very naturally wedded to the methodology of naturalistic inquiry. There is no setting of antecedent parameters and no predeter-

mination of output modes. The evaluator is guided mainly by whatever the audiences want to know (although he may insert some things they would rather not know), and that requires that he interact with them in their naturalistic settings. He responds to those issues with a variety of information, much of which will also be determinable only through face-to-face interaction. Throughout, the evaluator tends toward an evaluation report which may or may not be written but which endeavors to portray the state of affairs to the audiences. Value differences are taken into account (and indeed, are reflected from one audience to another so that their members are also aware of these differences in value orientations). The enterprise as described by Stake seems custom-made for naturalistic inquiry.

2. *The Judicial Model.*[21] The judicial model of evaluation utilizes the processes of the administrative hearing in the law as a paradigm. The purpose of such an evaluation, according to its chief proponent, Robert Wolf, is to illuminate, inform, and adjudicate *issues* based on the concerns of individuals affected by or involved with the object or activity being evaluated. This formulation is quite similar to that of Stake's responsive evaluation. The model differs from Stake's in the way issues are dealt with after they are identified.

Advocates or counsels who will take opposite views with respect to an issue are appointed by the sponsor of the evaluation. The role of these advocates is similar to that of real attorneys: they are to argue with as much vehemence, energy, and conviction as they can about the merits of the case. Supposedly, the "truth" about issues is more likely to emerge under such adversary conditions than with a singular evaluator making strenuous efforts toward complete "objectivity."

The two counsels go through a number of distinctive stages or steps in carrying out the evaluation:

(1) *Issue generation.* The issues are identified through "fact-finding interviews" with samples of the audiences involved, as in the case of the Stake responsive model.

(2) *Issue selection.* The purpose of this stage is to delimit the number of issues and to prioritize them, so that they may be manageable in a hearing format.

(3) *Preparation of formal arguments.* Each counsel or advocate team prepares formal arguments related to the selected issues. Available evaluation or other data may be used (to be introduced as "exhibits" in the hearing stage), and additional evidence may be collected, particularly evidence in the form of depositions from witnesses. Addi-

[21] The major references for the judicial model are Wolf (1975), Owens (1973), and Levine (1974).

tionally, selected witnesses may be asked to give testimony at the hearing itself.

(4) *Pre-hearing discovery sessions.* Each advocate team reviews the major arguments it intends to make and discloses the main features of its "evidence" for the other. Since the hearing is not a "trial" in the conventional sense, but an effort to determine "truth" as precisely as possible, each side shares its findings with the other so that the hearing may be as comprehensive as possible. In addition, the advocate teams decide on ground rules, e.g., number of witnesses to be called and criteria for determining admissability of evidence.

(5) *The hearing.* Modeled on an actual courtroom process, the hearing involves an administrative officer and a "jury" or hearing panel. After hearing the evidence, the jury carries out whatever tasks the advocate teams previously agreed to assign to it, which usually involve at least the determination of findings (which may include judgments of worth) and the making of selected recommendations.

Obviously, the judicial model is congenial to the use of naturalistic inquiry as a tool or technique. This is so especially in the issue-generation stage and the preparation-of-formal-arguments stage, since in both cases the advocates must go to involved audiences and participants to obtain their views. Furthermore, the form of the inquiry depends on the feedback received from these external sources and not on *a priori* conceptions held by the advocates.

3. *The Transactional Model.*[22] Robert Rippey suggests that his proposed transactional evaluation (TE) model is an alternative for more conventional evaluation, in that it examines not only the effects of changed programs but also the effects of these changes on the changers themselves and on the system as a whole. Thus, evaluation is a strategy for coping with change.

The new model is purportedly based on "open systems theory," which is appropriate to schools attempting more than a mere transmission of past traditions. An adaptive system has, at its heart, institutional research and development functions. Both adaptive and maintenance subsystems must be maintained within an organization. Ordinary formative evaluation is concerned only with assessing factors such as cost, time, and effort, and improving on them as much as possible. Transactional evaluation in addition recognizes the institutional disruptions brought about by attempts at change, taking them into account and attempting to alleviate them. Thus, evaluation seems to be a strategy for conflict management.

Transactional evaluation has five phases (Talmadge, 1975):

22The major reference for the transactional model is Rippey (1973). It should be noted that some evaluation scholars reject transactionalism as an evaluation mode, preferring to label it a management model that uses evaluation as a major tool.

(1) *The initial phase.* Pre-existing unrest or some other troublesome situation exists. A meeting is set up of interested parties under the direction of a "neutral" evaluator working in a non-judgmental atmosphere.

(2) *Instrumentation phase.* During this phase, a "Transactional Evaluation Instrument" (TEI) is developed whose purpose is to provide the evaluator with insight into the perceptions and expectations of various interest groups. The instrument also provides a forum for the sharing of opinions among the groups. The TEI is developed and administered in group sessions, during which (a) the evaluator initially formulates issues on the basis of general expressions from the group, (b) participants are asked to re-express the formulated issues in their own terms and express opinions about them, (c) the most representative and divergent of the written responses are carefully worded into items that can be rated on a scale from "strongly agree" to "strongly disagree," (d) the instrument is administered to the group, and (e) responses are examined.

(3) *Program development.* The program is redefined to reflect those goals and values on which the group can achieve some consensus.

(4) *Program monitoring.* Various groups agree to assume responsibility for implementing and monitoring the developed program.

(5) *Recycling.* As new conflicts emerge, the entire process is recycled to whatever phase is appropriate.

While the transactional model is not as naturalistically oriented as the preceding two, it does share certain features with them. It depends upon the identification of the problems or "trouble spots" that are causing difficulty. These problems exist in the minds of the people involved and can be identified only through some process of interaction with them. When the instrument is prepared, items are cast into the wording which is natural to the respondents, being taken almost entirely from their written opinion statements. Consensus is achieved through an interface process between the evaluator and the audiences, in which issues can be played back (and recycled) as often as necessary to achieve consensus.

The model does not depend upon arrangements of either antecedent conditions or output variables by the evaluator. Thus basic conditions for naturalistic inquiry are fulfilled. The procedures of responsive evaluation would be very similar to those of naturalistic inquiry if the process were loosened up a bit, going beyond the group meeting to permit a fuller exploration of people's opinions and perceptions of the issues.

4. *The Connoisseurship Model.*[23] This model, proposed by Elliot Eisner,

[23]The major references to the Connoisseurship Model are Eisner (1975) and Eisner (in press).

considers educational evaluation equivalent to educational criticism, i.e., the process through which the complexities of the entity being evaluated are penetratingly and publicly described and appraised. Criticism, in turn, depends on connoisseurship, the private act of appreciation, i.e., awareness of characteristics and qualities, of any set of phenomena. The analogy is based on artistic connoisseurship and criticism, and is derived from Eisner's extensive experience as an art educator.

The advance organizers of the connoisseurship model are "critical guideposts," i.e., values and concepts formed from tradition and habit as well as from explicit and implicit theories about the nature of standards for appraisal. These critical guideposts become known to and are internalized by the connoisseur through his training and experience. Connoisseurs are persons with refined perceptual apparatus, knowledge of what to look for, and a backlog of previous relevant experience. They have the ability to recognize skills, form, and imagination and to perceive the intentions and leading conceptions underlying the entity being evaluated. In effect, because of these characteristics, the connoisseur is himself the evaluation *instrument*. Having made his judgments, he communicates the qualities that constitute the entity being evaluated, its significance, and the quality of experience engendered by interaction with it, often through the use of rich metaphors.

This model also has many elements which relate it to naturalistic inquiry. The connoisseur approaches the entity to be evaluated with a general framework of critical guideposts, but he does not structure his stimulus in any way. The outputs of that interaction are also variable, and depend in part on how deeply the connoisseur/critic wishes to delve. Different critics may, with equal validity, apply different critical guideposts and therefore provide different descriptions and judgments. Each formulation constitutes a valid way of viewing the object. The connoisseur has a great deal in common with naturalistic inquirers (consider anthropologists, for example) who consider themselves as instruments who, by virtue of training and experience, know how to enter a situation and produce an appropriate description/judgment about it.

5. *The Illumination Model.*[24] The illumination model proposed by Malcolm Parlett and David Hamilton might be thought of as a special form of the earlier-discussed responsive model. It seems to provide an additional element to evaluation beyond merely adjudicating worth, namely, the illumination of problems and issues.

Its relation to naturalistic inquiry is signaled early by the authors in the distinction they draw between the "agricultural-botany" paradigm, which they assert uses a hypothetico-deductive methodology derived from the

[24]The major reference for the illumination model is Parlett and Hamilton, (1977).

experimental and mental-testing traditions in psychology, and an emergent "social anthropology" paradigm that relates to social anthropology, psychiatry, and participant observation research in sociology. The latter paradigm is the basis for illuminative evaluation:

> The model described here, *illuminative evaluation*, takes account of the wider contexts in which education programs function. Its primary concern is with description and interpretation rather than measurement and prediction. It stands unambiguously within the alternative methodological paradigm. The aims of illuminative evaluation are to study the innovatory program: how it operates; how it is influenced by the various school situations in which it is applied; what those directly concerned regard as its advantages and disadvantages; and how students' intellectual tasks and academic experiences are most affected. It aims to discover and document what it is like to be participating in the scheme, whether as teacher or pupil, and, in addition, to discern and discuss the innovation's most significant features, recurrent concomitants, and critical processes. In short, it seeks to address and to illuminate a complex array of questions. (Parlett & Hamilton, 1977, p.144)

Illuminative evaluation is carried out in three stages: (1) initial observations for the purpose of familiarization with the day-to-day reality of the setting(s), largely in the manner of social anthropologists or natural historians; (2) more sustained and intensive inquiry into a number of common incidents, recurring trends, and issues frequently raised in discussion; and (3) efforts to seek general principles underlying the organization of the program, determine patterns of cause and effect within its operation, and place individual findings within a broader explanatory context. "Beginning with an extensive data base, the researchers systematically reduce the breadth of their enquiry to give more concentrated attention to the emerging issues. This 'progressive focussing' permits unique and unpredicted phenomena to be given due weight" (Parlett & Hamilton, 1977, p.148).

The methods used in illuminative evaluation include observation, interviews, questionnaires and tests, and documentary and background sources. Observations are intended primarily to build up a continuous record of ongoing events, to add interpretive comments on manifest and latent features of the situation, and to uncover tacit assumptions, interpersonal relationships, and status differentials. Interviews are used primarily to determine the perceptions and views of individual participants. Questionnaires and tests are included to permit reaching of wider audiences and to obtain information that sustains or qualifies earlier, tentative findings. Documentary and background sources provide information about the history of events.

Evaluation services decision-making but is concerned more with the collection of relevant information than with the act of decision-making itself:

> Illuminative evaluation thus concentrates on the information gathering rather than the decision-making component of evaluation. The task is to provide a comprehensive understanding of the complex reality (or realities) surrounding the program—in short, to "illuminate." In his report, therefore, the evaluator aims to sharpen discussion, disentangle complexities, isolate the significant from the trivial, and to raise the level of sophistication of debate. (Parlett & Hamilton, 1977, p. 153)

Illuminative evaluation thus seems to share many of the purposes of responsive evaluation but appears to be even more committed to naturalistic approaches.

These five models, responsive, judicial, transactional, connoisseurship, and illuminative, are thus seen to have close philosophic and operational ties with naturalistic inquiry. Their emergence at this time argues strongly for the utility of naturalistic inquiry for the field of educational evaluation, and helps make the case that N/I should be acceptable as an alternative evaluation methodology.

THE FIT BETWEEN EVALUATION'S MAIN THRUSTS AND THE STRENGTHS OF N/I

Despite their other differences, most scholars working in the field agree that evaluation has two particular thrusts or elements: description and judgment. Evaluation is certainly not complete without a description of the entity being evaluated, its setting, the way in which it is viewed by various participants, its processes, its outcomes. But evaluation is similarly incomplete if it does not also comment on the worth, utility, or merit of the evaluated entity as well. Indeed, the root term of the word, "evaluate," indicates the premium placed on this valuing process.

Both these elements are particularly suited to the methods of naturalistic inquiry. Description is indeed the basic forte of that approach—to determine what is there, to discover how components are related, to determine what people think about a situation. All of the emergent models discussed in the preceding section lean heavily on description of this sort as the basis for their unique approaches.

The problem of valuing—of determining merit, utility, or worth—obviously cannot be dealt with by the methods of science. Values are intensely personal, social, and cultural. What is valued by one individual or group may be devalued by another individual or group. Most of the evaluation approaches referenced above rely on some concept like issue, problem, concern, or question as the basis for analysis, and such concepts imply an underlying value structure. N/I appears to provide a methodology well suited for uncovering, interpreting, and reporting such elements in a most useful way.

Chapter IV:
SELECTED METHODOLOGICAL PROBLEMS OF NATURALISTIC INQUIRY

There are many methodological questions that can be raised about naturalistic inquiry, ranging from basic epistemological issues to operational or procedural matters. As noted in the introductory section of the monograph, attention here will be limited to some "middle range" problems that fall between these two extremes. For convenience these problems will be classified as *boundary* problems, *focussing* problems, and problems of *authenticity*. These three classes will be briefly described in this section; the subsequent three sections of this chapter will deal with each class in some detail.

Boundary Problems

The boundary referred to is the boundary of the inquiry as a whole. It was noted in Figure 2 and its accompanying discussion that one feature distinguishing naturalistic from other forms of inquiry is the extent to which the investigator places constraints on antecedent conditions. In experimental inquiry, these conditions are sharply constrained (controlled), so that there is no question as to the boundary of the inquiry. But in the case of naturalistic inquiry, antecedent conditions are not constrained in any way; presumably any antecedent condition may become a focal point for the study. Thus, the boundary problem is simply this: How is the inquirer to set limits to his inquiry? What are the rules for inclusion and exclusion? How can the inquirer determine what is and is not relevant?

The argument will be made that, while the naturalistic investigator does not impose *a priori* constraints, neither does he approach his task in a mindless fashion. Having an open mind is not equivalent to having an empty one. In general, the N/I inquirer has some problem to investigate (as do all inquirers), and the parameters of that problem serve to determine the inquiry limits. In the area of evaluation specifically, the boundaries are set partly by the initial stipulations made by the evaluator's client and/or sponsor (a surrogate problem) and partly through a recycling process common to all naturalistic inquiry that helps him to converge on the final boundaries.

Focussing Problems

Figure 2 also illustrates the point that naturalistic inquirers seek to avoid imposing constraints on outputs insofar as possible. If outputs are not to be

defined before the fact, then whatever outputs do occur must be collected, analyzed, categorized, and interpreted after the fact. This situation gives rise to two sub-classes of problems which will be termed the problems of *convergence* and *divergence*.

1. *Convergence.* Somehow the naturalistic inquirer must derive a set of units or categories within which he will classify and interpret observed outputs. A number of questions arise:

a. What is the *nature* of a category? How can the inquirer identify a category when it emerges in an interaction with an information source, (i.e., an informant, an observed subject, a document)? What inferences may reasonably be drawn by the inquirer about *the basis* for a category that seems to exist in a respondent's mind or that apparently underlies his behavior?
b. What is the basis for *unitizing, categorizing,* and *prioritizing* information? What considerations enter into establishing categories? What guidelines exist for the process of classifying particular observations? How can the relative importance or salience of derived categories be established?
c. What are the *criteria* for determining when a necessary and sufficient set of categories has been derived? That is, what are the criteria for determining when the categorizing process has gone far enough?

2. *Divergence.* Once the naturalistic inquirer has a "fix" on the situation by virtue of having derived a set of categories for dealing with it, he is usually interested in "fleshing out" the categories with additional information required for completeness and thoroughness. From the focal point of the category itself, the inquiry fans out or diverges to include as many data items and perspectives as are relevant. Again, several questions emerge:

a. What *strategies* exist which the naturalistic inquirer might employ to undercover relevant information?
b. What is the basis for *inclusion or exclusion* of any particular datum?
c. When should the information collection *stop?* How can the inquirer be sure that he has adequately "covered" (or "uncovered") the domain of his interest?

Problems of Authenticity

The major purpose of this monograph is to propose a methodological alternative to conventional inquiry for educational studies in general and evaluation studies in particular. The point was made early

- that the idea of "method" evolved as a safeguard against error;

- that no knowledge can ever be secure from error, but that established and accepted methods can provide relative peace of mind on that point;
- that if a given method is to be taken seriously as an alternative to existing methods, bases for trust in *it* will have to be established.

The issue of "authenticity"[25] is concerned with the establishment of such bases. How might the naturalistic inquirer argue to persuade a methodologically sophisticated peer about the authenticity of his information, the categories he derives from it, and the interpretations he makes of it?

It will be suggested that the three major criteria typically applied to social inquiry (validity, reliability, and objectivity) are in fact also applicable to naturalistic inquiry, but in a reinterpreted sense. Several new terms will be offered that seem to reflect these reinterpretations more adequately than does conventional terminology.

BOUNDARY PROBLEMS

Naturalistic inquiries are not mounted *in vacuo*; like all other forms of inquiry, they emerge in response to a perceived problem. This fact is of great importance, for it suggests a means by which the investigator can limit his inquiry without imposing constraints that would force it into unnaturalistic forms. Before assaying a description of that process, however, it will be useful to consider briefly the nature of problems and the forms that they may take.

Problems do not exist in nature but in the *minds of people*, a crucial fact and one of the main reasons why one might recommend the naturalistic method in the first place. Despite the centrality of the concept of "problem" in all inquiry, the term remains peculiarly unexplicated. Research texts devote large sections to procedural issues but treat the nature of problems very cavalierly, often defining a "problem" simply as a question or an objective ("The problem of this research is to . . ."). A proper understanding of this pervasive concept is a first step toward integrated, directed, and meaningful inquiry, whether pursued by naturalistic or other means.

The definition of the term "problem" is of course arbitrary; what is needed is a definition which is maximally heuristic. "Problem" is defined here as follows:

> A problem is a situation resulting from the interaction or juxtaposition of *two or more factors* (e.g., givens, constraints, conditions, desires, etc.) which yields: (1) a perplexing or enigmatic state (a *conceptual* problem); (2) a conflict which renders the choice from among alternative

[25] The term "authenticity" is introduced not so much in the interest of contributing to a vocabulary unique to N/I (and thereby perhaps enlarging the mystique of the method) but for the sake of avoiding connotations of conventional terms that are inappropriate.

courses of action moot (an *action* problem); or (3) an undesirable consequence (a *value* problem).[26]

It is convenient to think of problem statements as similar in form to the logical syllogism; the two juxtaposed or interacting factors are akin to the syllogistic propositions, while the formal problem is akin to the syllogistic conclusion. The following are examples of each of the three types of problems when cast into the form of syllogisms:

1. *Conceptual problem* (a perplexing or enigmatic state). In this type of problem, the two juxtaposed elements are logically, conceptually, or theoretically inconsistent, e.g.:[27]

> *Statement 1*: The excellent state of affairs seems to emerge most surely when those on whom it depends feel that they are not being judged and evaluated, feel that their failures and shortcomings are understood and accepted, feel that their worth—as teachers, as administrators, as professionals—is unconditional, that is, not contingent on how they score on anyone else's scale of merit.
>
> *Statement 2*: And yet it appears that people move truer and more certainly toward excellence to the extent that they clarify their purposes, measure the impact of their action, judge it, and move on—in a few words, evaluate their progress.
>
> *Conclusion (the problem)*: Evaluation seems to pose the paradox that, while its application helps people chart a truer course toward excellence, it may be dysfunctional to their performance as human beings.

The evaluator is confronted by a dilemma. Can both statements be true? Is there no escape? Can an evaluation situation be structured so as to avoid the personal dysfunctionality while still enhancing the organizational benefits? Can evaluation be conceptualized in a different way so as to minimize the worst consequences of this problem (as Rippey has tried to do in using evaluation as a conflict management tool)? A dozen or more approaches that an evaluation researcher might use in inquiring into this problem can easily be imagined, including both naturalistic and conventional methods.

2. *Action problem* (a conflict which renders the choice from among alternative courses of action moot). In this type of problem, the two juxtaposed elements call for two apparently mutually exclusive actions, e.g.:

> *Statement 1*: Schools of education are being called upon to respond to problems and situations (e.g., multicultural education, integration, in-service re-education of teachers, field-based training) that move them further from the academic setting and more closely into actual school sites.

[26]This definition and much of the following discussion is based upon unpublished materials that my colleague at Indiana University, Professor David L. Clark, and I developed over the past 13 years in connection with the inquiry courses we have taught at Indiana and Ohio State Universities.

[27]The two statements comprising this example are taken from Glass (1975).

Statement 2: University promotion and tenure systems, under the goad of decreasing budgets, declining enrollment and the existence of large proportions of faculty already tenured or appointed at upper professorial ranks, are stressing traditional criteria of research and scholarly activity.

Conclusion (the problem): Individual professors in schools of education are faced with a Hobson's choice: fulfill institutional responsibilities while risking personal academic careers, or fulfill personal career needs while exposing their school to charges of negligence and ineffectiveness.

Again, an inquirer senses a dilemma worthy of investigation. What are the consequences of each of the choices implied in the "conclusion" statement? What proportion of faculty is susceptible to this problem? What is their perception of the "fairness" of the situation? Are university administrators outside schools of education aware of the conflict and its possible disastrous outcomes? For professors, what personal characteristics make it more or less possible to live with this problem?

3. *Value problem* (an undesirable consequence). In this type of problem, the two juxtaposed elements lead to an outcome that is undesirable or inimical in terms of some value standard, e.g.:

Statement 1: Testing of children often leads to the labeling of some of them as "handicapped" and to their assignment to special educational programs.

Statement 2: The milieu of special school classes is often very different from that of a normal school environment, particularly as it calls the attention of the child to the fact that he is somehow "different."

Conclusion (the problem): Handicapped children are reinforced in their perception of abnormality and lose their motivation to perform at "normal" levels.

Again, a number of questions surge into the investigator's mind. Can testing be carried out under circumstances that do not lead to labeling? Is mainstreaming possible as a response to this condition. That is, can appropriate remedial or therapeutic measures be undertaken in the "regular" classroom? What teacher retraining must occur and with what materials? Can the children be given some form of therapy that would help them overcome these negative self-perceptions? What is the nature of their perceptions?

Boundaries Suggested by the Syllogistic Problem Format

If we accept this formulation, it is apparent that the naturalistic inquirer, once he has placed his problem into the syllogistic format, will always have several strategies immediately available in determining the boundaries of his inquiry:

1. He can document the "facts" asserted in the propositional statements. The problems exist only to the extent that the propositions are valid. In the examples above we have accepted them at face value, but questions can certainly be raised. Is it true the people are productive only to the extent to which they feel that they are worthy on their own terms? Is it true that university reward systems have tended to emphasize traditional criteria of research and scholarly productivity? Is it true that the milieu of special classes is so different from that of regular classes? This *verification* or *documentation* mode places obvious bounds on the inquiry.

2. He can look for the causes of these states of affairs. Why do "people move truer and more certainly toward excellence to the extent to which they clarify their purposes?" Why do motivational levels drop in children when they are placed in special classes? Why do institutional administrators perceive a tightening of promotion and tenure requirements as partial solutions to their fiscal dilemmas? The *causality* mode of inquiry also places obvious bounds on the domain of investigation.[28]

3. He can look for mediating or ameliorating factors that might serve to contravene the propositions. Could evaluation become more humanistic in its orientation?[29] Could university administrators and faculty members be persuaded to apply different criteria in professional school promotion/tenure cases, as they do, for example, in relation to fine arts departments by accepting art work in lieu of more conventional scholarly products? Could overt labeling of children as "handicapped" be avoided? This *contraventional* mode of inquiry (which might be thought of as seeking *solutions* to problems) also sets the stage for an inquiry type.

4. He can determine the consequences if the propositions are permitted to go unchallenged or not ameliorated. The consequences are of two kinds: real and perceived. What, in fact, is the effect of performing evaluations without regard to their possibly dysfunctional consequences? What do various relevant groups perceive as consequences? How do these perceptions relate to actual consequences? What are the results for the behavior of various groups of perceivers, regardless of whether their perceptions are valid? Similar questions can be asked about the consequences of a conflicting reward system or special class placement of handicapped youngsters. This *consequential* mode of inquiry is another means for setting inquiry boundaries.

There are probably other classes of strategies for inquiry that would serve to set reasonable bounds on any investigation. However, the four strategies

[28]The determination of causality has long been defined as the special province of experimental inquiry. But Scriven's proposed "modus operandi" method is an alternative entirely open to the naturalistic investigator (Scriven, 1974).

[29]Glass poses this alternative but dismisses it as "too gentle."

just described will adequately make the case. Naturalistic investigators need not feel at a loss for means to set reasonable bounds to their inquiries. Problems can be stated explicitly in ways that make verification, causality, contraventional and consequential approaches definable and, hence, bounded.

The Special Case of Evaluation

The previous discussion clarifies the fact that any inquiry can be bounded through the use of the problem statement as a limiting tool. Evaluation can be conceptualized as responding to problems as well and so, be bounded; but there are two special circumstances that exist in evaluation inquiries which will also markedly determine the boundaries of the study:

1. The problem to be studied is not, initially at least, specified by the evaluator but by his client or his sponsor.[30] There seems to be no legitimate way to avoid this situation, nor indeed should it be avoided. The client is entitled to his interpretation of the situation and to issue those "marching orders" that he believes are appropriate. But the evaluator need not be naive. He must recognize that the client may have many covert reasons for putting the charge as he does, including, for example, selecting for evaluation only those program aspects that appear to be successful, covering up program failure by focusing on partisan testimonials, making evaluation gestures designed to promote a favorable image, responding to governmental mandates which are not taken seriously but must be overtly complied with, and the like. Sheer ignorance may also shape the client's directive.

The evaluator needs to be aware of these pitfalls and to guard against them. The time to negotiate these potential problems is at the beginning of the evaluation and not at the end. For example, Stufflebeam's "Administrative Checklist for Reviewing Evaluation Plans" is very helpful in negotiating such differences. (Stufflebeam, 1974) The Joint Committee on Evaluation Guidelines and Standards is currently developing other guidelines that will also be useful. (Joint Committee, 1977).[31]

---

[30] A traditional criterion for distinguishing between applied research and "mere" service is the source of the problem. If the problem is specified by the investigator, it is an applied study; but if the problem is defined by the client, it is service. Many clients are frustrated to find consultants reshaping their problems into the consultant's terms in order to serve the consultant's need to believe (and purport) that his study is genuine research. ("You've got an interesting situation here, but you're asking the wrong questions.")

[31] The Joint Committee is a national project sponsored by the American Educational Research Association, the American Psychological Association, the National Council for Measurement in Education, and a number of practitioner-oriented professional associations. Currently funded by grants from the Lilly Foundation, the National Science Foundation, and the National Institute of Education, it is chaired by Daniel L. Stufflebeam of Western Michigan University. It has devised a set of guidelines and standards in draft form which have been tested with a national panel of reactors and soon will be critiqued in a series of national hearings to be scheduled by the Committee.

The naturalistic evaluator has another way out of the constraints that might otherwise be placed on him by the client's definition of the problem—the recycling process that he would normally go through. We have noted in an earlier section that naturalistic inquiry typically proceeds in a "wave" form and leads the investigator to *cycle* through a series of observations alternately directed at discovery and verification. Thus, the N/I investigator views the initial formulation by the client as simply one stage of discovery which leads naturally to another point in the "wave"—that of verification. If the client has deliberately or inadvertently distorted the situation, or given a charge which involves covert motivations, the N/I investigator is quite likely to uncover these facts and be able to take them into account. He may, in fact, reflect back to the client the inconsistencies between his mandate and the situation as he finds it. (The nature and extent of this reflection obviously depend heavily upon the politics of the situation and the quality of the relationship between evaluator and client.) There is no valid reason to suppose that the N/I evaluator will simply be the dupe of the client, if he follows the normal strategies of naturalistic inquiry.

2. The fact that evaluations are in general identified as such to everyone creates special relationships between the evaluator and the other parties. Boundaries are set *de facto*. A role is set both for the evaluator and for all other persons with whom he comes in contact. Fairness becomes a crucial element, and certain legalities and sanctions are invoked, e.g., the right to privacy. (More will be said about fairness as an element in judging an evaluation later.)

The entity to be evaluated is also known to all concerned. This fact poses certain additional constraints. Subjects are likely to take a jaundiced view of questions that, in their opinion, range beyond the legitimate area of inquiry.

There is a caveat, however: the evaluator cannot take for granted that a subject's definition of legitimate bounds is appropriate. But surely the normal procedures and tactics of naturalistic inquiry make it unlikely that the N/I evaluator will permit himself to be inappropriately boxed in, or fail to uncover relevant information barring a massive conspiracy on the part of respondents to keep it from him.

FOCUSSING PROBLEMS

Focussing problems have been defined as emerging from the analysis, categorization, and interpretation of *outputs* occurring in the natural situation. Two sub-categories of problems were identified: problems of *convergence*, involving the development of categories within which data may be assimilated, and problems of *divergence*, involving the "fleshing out" of categories with whatever additional information is required for completeness and thoroughness. Each of these sub-problems will be discussed below.

Problems of Convergence

Problems of convergence arise because the naturalistic inquirer must derive a set of units or categories within which he will classify and interpret observed outputs. In the specific case of evaluation, a rubric exists which greatly simplifies the naturalistic evaluator's task.

The discussion of various emergent models of evaluation noted that the concept of "issue" or "concern" permeated a number of them. Thus, the responsive model of evaluation responds to audience concerns; the judicial model deals with issues and concerns raised by individuals affected by the entity being evaluated; the transactional model formulates issues by defining current unrest or troublesome situations; the illumination model "progressively focuses" on issues. Thus the concepts of issues and concerns appear to be useful. *It is here proposed that the categories of an educational evaluation when conducted naturalistically should be such concerns and issues—that concerns and issues should be the "advance organizer" for the evaluator's efforts.*[32]

Issues and concerns do not exist in nature *but in the minds of people.* The naturalistic evaluator's task is to identify the people whose issues and concerns can or should be taken into account, whether or not the involved persons are themselves willing to press their case. He must devise means to determine what those concerns and issues are.

A distinction needs to be drawn between "concern" and "issue," since the two terms are not precisely identical. Stake defines an *issue* as

> . . . a circumstance about which people disagree. It usually involves a condition having some features causing (or believed to cause) certain effects. These effects are valued differently by different people—so they disagree as to whether and how the condition should be changed.
>
> The ingredients for an issue, then, are the condition, the effects, the relationships between condition and effect, the different valuing, and the alternatives among courses-of-action for changing conditions.
>
> (It is true that the contention might be due more to disagreement as to whether or not a relationship holds than to different valuing of the effects. Either way, issues are points of contention.)[33]

The phrase, "points of contention," seems to be a particularly felicitious way of expressing the essence of an issue. Note also that contention requires at least two parties—two individuals or two groups. Issues cannot (except in the form of hearsay) be extracted from a single source.

---

[32]It is important to note, however, that N/I need not be confined to evaluation modes based on concerns/issues as the advance organizers. N/I could also be used to investigate the achievement of objectives, servicing of decisions, determination of effects, etc.

[33]Quoted from an internal document prepared by Robert E. Stake for the "Case Studies of Science Education" project. See p. 48.

Concerns, on the other hand, may have a singular referrent. A concern is a matter of interest or importance to its holder. Concerns generally have a future orientation, with overtones of fear or anxiety. In an actuarial sense, concerns may grow in importance as more and more individuals or groups share them, but even singular concerns may be of great meaning and cannot be ignored.

Given this background, we may venture the following definitions:

- An issue is a point of contention between two or more parties.
- A concern is a matter of interest or importance about which one or more parties feel threatened.

Sources of issues and concerns. The naturalistic evaluator, in probing for the concerns and issues that characterize a given situation, is materially assisted by keeping in mind that there are certain standard situations in which the persons involved may see the entity being evaluated as giving rise to issues or concerns. Among the more important sources of these issues or concerns are the following:

1. *Undesirable consequences of an interaction.* The institution of special classes, for example, may produce a loss of self-esteem in students assigned to them as they come to realize the "special nature" of the group.
2. *Confusion regarding courses of action.* The installation of a new administrative format for reporting, for instance, that is based on functional rather than organizational lines may cloud rather than clarify lines of authority.
3. *An undesirable deviation from older practice.* This might be called the "What was good enough for my Daddy" syndrome. Thus, parents may reject "new math" approaches on the grounds that they do not teach the computational skills needed for balancing checkbooks or computing most favorable prices on grocery items—skills they were taught and found useful.
4. *Conflicts with traditional values.* Thus, blacks may view schools operated by whites as teaching values inimical to their cultural heritage.
5. *Conflicts with personal values.* A person with Amish values may see teaching biology as a threat to his value system.
6. *Potential loss of power.* The institution of an evaluation office may be perceived as providing too much information to the wrong people, who can then use it to attack entrenched interests.
7. *Potential economic threats.* The institution of a system of teacher aides may funnel funds into the hands of selected poor and away from more vested interests.

8. *Perceived inconsistency with a suggested course of action.* A parent who has suggested the need for a disciplinarian is not pleased with the appointment of a social worker as a vice-principal.
9. *Lack of understanding of rationales or goals.* Many parents reject the rationale underlying busing as a means to overcome racial prejudice.
10. *Bias based on a negative personal experience with a given institution.* Parents raised during the period of "life adjustment" education, popular in the schools of the 50s, may be somewhat suspicious of school programs focussing on cultural heritage rather than on basic skills.
11. *Potentially harmful side effects.* Parents often object to materials used in sex education courses because they fear they will be used as how-to-do-it manuals rather than as moral deterrents.

It should not be assumed that all issues and concerns are negative. Positive issues or concerns may exist because persons wish to determine ways in which an entity may be improved (much formative evaluation would fall into this category); because they wish to test an entity in which they already have a great deal of confidence or faith; because they wish to produce information showing that the entity being evaluated is at least as effective as its competitors; or because they wish to respond to real or imagined criticism from others. No doubt other reasons could be provided.

Nor should it be assumed that this list is intended as an *a priori* set of categories which every naturalistic evaluator would be well-advised to use. Concerns and issues are unique to each situation. The general list provided above is intended to give some insight into the nature of the motivations that provoke concerns and issues; as such, they are useful cues for listening and probing.

It seems likely that all concerns and issues could be classified within some general taxonomy. At some future time it may be productive to attempt the development of such a classification system. Surely these elements would have to find a place in that system:

1. Prudential factors.
   a. Economic factors (chiefly costs).
   b. Resource allocation factors.
   c. Trade-offs with other options (including doing nothing and/or "cheapie" versions.)
2. Moral/ethical factors.
3. Political factors.
4. Human factors.
5. Legal factors.
6. Scientific factors.

7. Value factors.
   a. Personal values.
   b. Socio-cultural values.

Again, this list is not offered as definitive but as suggestive of what might emerge.

Unitizing/categorizing processes. The task of converting field notes and observations about issues and concerns into systematic categories is a difficult one. No infallible procedure exists for performing it. Since this process is, to such a major extent, "arty" and intuitive, it is often viewed as "merely" subjective and, hence, not likely to lead to results worthy of serious consideration. There are, however, several steps which the naturalistic inquirer can take that may be useful.

A first step for the inquirer is to look for *recurring regularities* in his sources. Do the same kinds of observations or comments recur at different times, from different informants or documents, in somewhat different contexts? These regularities form the basis for an initial sorting of information into categories that will ultimately be labeled as concerns and issues.

In this process it is likely that a considerable number of individual data items will be placed into an "other" or "miscellaneous" group because they do not seem to fit. In some cases these data items can safely be ignored as subjective, idiosyncratic observations of no real substance. On the other hand, it may well be that some represent incisive observations or judgments made by persons uniquely competent to make them. Often the evaluator will not be able to tell whether a given observation is of the former or latter variety. His tendency ought to be to take everything seriously until he accumulates evidence to the contrary; the importance or salience of an item need not be a function of its frequency of notation.

Once a preliminary set of categories has been developed, the naturalistic evaluator will run certain systematic checks. The utility of his category set is largely a function of the *internal homogeneity* among items classified in any particular category and the *external heterogeneity* among categories. If a given category is to be defensible as encompassing a single concept, all of the items within it ought to exhibit high homogeneity. Are they logically related? Do formulations of a particular issue or concern from different perspectives (e.g., those of a protagonist and of an antagonist) appear congruent in the way the issue or concern is defined? Do examples and counter-examples "dovetail" in a meaningful way? Similarly, differences among categories ought to be bold and clear. The existence of a large number of unassignable or overlapping data items is good evidence of some basic fault in the category system.

The naturalistic evaluator will realize that even a well-refined category system based upon an initial wave of observations and interviews is likely to be inadequate. Referring back to the "wave" concept as outlined in Figure 1, we may imagine that such an initial formulation represents a "discovery" phase which will then lead immediately to a "verification" phase. During this latter phase the N/I evaluator can test the utility, scope, and inclusiveness of his category system. Both alterations and additions are usually necessary. This identification of gaps leads into the second "discovery" phase, and so on. Thus, the category system is a dynamic entity, constantly changing and, hopefully, improving. Collection, coding, and analysis of data go on concurrently as an interactive process, "unfolding, building, and contouring itself to the reality that exists in peoples' experiences and perceptions." (Wolf & Tymitz, *in press*, p.2)

There are several caveats which the naturalistic evaluator should bear in mind in this process:

1. He probably cannot hope to devise as "exhaustive" a set as he might like. Concerns and issues are probably endless and, even after prioritizing them, he will not be able to convert all of his observations into an easily manageable set (short of using a residual category into which to place large numbers of items).

2. Nor is it likely that he will ever be able to devise a scalar set in which categories are ordered in terms of magnitude or some other identifiable dimension. He will need to be satisfied with discrete categories that have no ordinality.

3. The naturalistic evaluator can and should, however, strive for categories that are unidimensional (internal homogeneity). It should be remembered that, as a next step after categories are initially defined, the N/I evaluator will go back to the field to "flesh out" his categories. That task will be enormously complicated if he has not exercised care in establishing his categories in the first place.

4. Finally, the evaluator should aim for an appropriate level of discourse for the categories on which he does settle. That level serves to determine the level at which issues and concerns are analyzed. Selection of too broad a level may cause the evaluator to overlook significant concerns and issues because they are "buried" within a larger category, while selection of too narrow a level may focus attention on detail to the exclusion of the "bigger picture."

Prioritizing processes. It is likely that the set of categories which the naturalistic evaluator will derive in any given situation will be too large to be manageable, especially in the next stage of the evaluation in which the categories are to be "fleshed out." Further, resources will typically not

permit the exploration of more than a limited set. What should guide the evaluator in prioritizing the categories which do emerge so that he can focus on those most worthy of further exploration? While hard and fast rules cannot be set down, it seems likely that the evaluator would be well served by attending to these considerations, at least:

1. *Salience or weightiness.* It has already been noted that the number of respondents who mention a particular issue or concern is not a sufficient criterion for assigning priority; nevertheless that index can be useful in the sense that issues and concerns more frequently identified have more salience in the situation and many therefore be more weighty with one or more audiences. Surely one would not wish to eliminate an issue or concern that received frequent mention.

2. *Credibility.* As attested by the recycling process which the N/I evaluator uses with his sources of information, certain items will simply be accorded more credibility by the several audiences; some items will strike them as highly realistic and others as more or less incredible. Since the audiences are probably more in touch with the reality of that situation than is the evaluator, their estimates of credibility should be highly determinative in what is retained.

3. *Uniqueness.* Some concerns and issues may stand out by virtue of their uniqueness; they are noticeable simply because they are so different from other items. Now of course their difference may be accounted for because they are the product of highly idiosyncratic perspectives. Assuming they have some credibility, however, unique items probably ought to receive higher priority simply because they add interesting detail and proportion to the evaluator's perspective.

4. *Heuristic value.* Some concerns and issues ought to be retained because they have the property of opening up inquiry areas not otherwise recognized, or because they provide a unique leverage on an otherwise common problem. There are always some ways of formulating problems that are more productive than others, even when the root problem is the same. Thus, some concerns and issues have special value in providing points of entree for the evaluator and/or in facilitating his inquiry.

5. *Feasibility.* Some concerns and issues, whatever their inherent interest, are simply not susceptible to inquiry. The methodology may not be available; there may not have been enough time for an effect to become noticeable; political factors may militate against exposure to public view; resources may simply be inadequate. Rejection of a concern or issue on feasibility grounds may at times be painful because of other characteristics that would seem to make that concern or issue an especially fruitful one to pursue, but may nevertheless be necessary.

6. *Special interests.* Some concerns or issues not intrinsically interesting may nevertheless be important to pursue because of the special interest taken in them by a powerful audience, e.g., project administrators, funders, or legislators. Clearly such matters cannot be ignored.

7. *Materiality.* Materiality may be the most important criterion to bring to bear; it is the most *substantively* oriented criterion included here. Concerns and issues that have an obvious and important bearing on the entity being evaluated should be included, even though at times that may not suit the purposes of relevant others and some audiences may, in fact, prefer *not* to have those issues and concerns aired. Materiality obviously cannot be the sole basis for prioritizing; to assert that it is would be to remain blind to the fiscal, social, and political realities that surround every evaluation. Nevertheless, materiality is of great import and must serve as the basis for the initial cut, as it were, with the previously stated criteria helping to make adjustments and refinements.

Determining the completeness of a set of categories. How can a naturalistic evaluator finally determine whether or not the categories he has articulated constitute a necessary and sufficient set? Allusion has already been made to the fact that there should be a minimum of unassignable data items and relative freedom from ambiguity of classification. But there are other considerations:

1. The set should have internal and external plausibility, a property that might be termed "integratability." Viewed internally, the individual categories should appear to be consistent; viewed externally, the set of categories should seem to comprise a whole picture. Often an incomplete set will suggest other categories for which no data exist. Such categories could not have emerged inductively from the data collected to that point but seem to be required by the logic of the sub-set of categories that does emerge. The N/I evaluator would take this situation as a stimulus to look for data that would corroborate the existence of the suggested categories.

2. The set should be reasonably inclusive of the data and information that do exist. This feature is partly tested by the absence of unassignable cases, but can be further tested by reference to the problem which the inquirer is investigating or by the mandate given the evaluator by his client/sponsor. If the set of categories does not appear to be sufficient, on logical grounds, to cover the facets of the problem or mandate, the set is probably incomplete.

3. The set should be reproducible by another competent judge. This injunction should not be taken too literally, however, in the sense of requiring that the set be reproducible from the raw data by a second inquirer who has no contact with the first. We have already noted the situation of

"multiple reality" (pp.15ff.). Different investigators may choose to peel the onion of reality to different levels, depending upon the purposes, interests, experience, and expertise that they bring to the task. Two completely independent observers thus could not be expected to devise precisely the same set of categories. The second observer ought to be able, however, to verify that (a) the categories make sense in view of the data which are available, and (b) the data have been appropriately arranged into the category system.

This process is similar to the audit process which is common in business and is being increasingly used in the field of evaluation. The fiscal auditor performs two functions: he attests that the *system* of accounting is appropriate (and honest), and that the *bottom line* is correct. In similar fashion the category system auditor may be called upon to attest that the category system "fits" the data and that the data have been properly "fitted into" it.

4. The set should be credible to the persons who provided the information which the set is presumed to assimilate. This property of credibility might also be termed "contextual appropriateness." Such credibility is far and away the most important criterion, particularly in view of the fact that what the set purports to summarize are issues and concerns in the minds of people. Who is in a better position to judge whether the categories appropriately reflect their issues and concerns than the people themselves? More will be said of such credibility checks in connection with the problems of authenticity to be discussed in the next section.

Problems of Divergence

When the naturalistic evaluator has identified even a preliminary set of categories he will wish to begin "fleshing" them out, i.e., by collecting information which will describe the issues or concerns in some detail, by providing perspectives for viewing them, and by developing sufficient evidence to permit judgments to be made about them.[34] How shall he go about that task? What are some strategies he can employ in searching for the needed information? How can he tell whether to include or exclude any particular datum? How can he tell when to stop collecting information?

Strategies. The typical situation confronting the N/I evaluator is shown in Figure 3. A focal area, i.e., an issue or concern, has been identified, whose boundary, depicted as a broken line, is not well known. Within that area, certain information has already been identified. This information, in fact,

[34]No attempt will be made here to deal with the difficult question of who should make the judgments: evaluator, decision-maker, audience, etc. Answers to that question depend heavily on the point of view of the respondent.

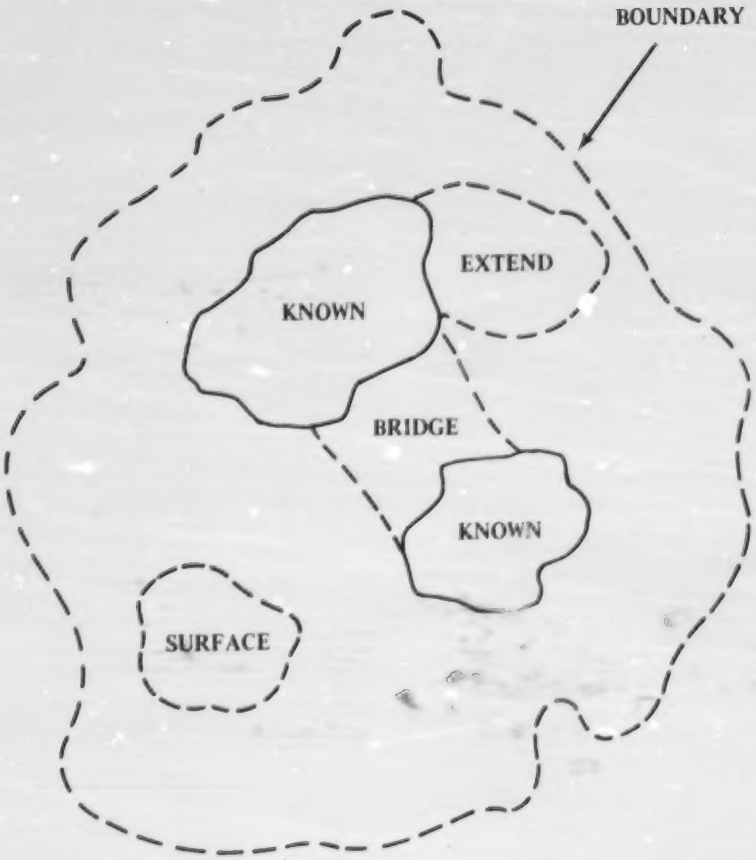

Figure 3: Representation of Divergence Strategies: Extending, Bridging, and Surfacing

served as the inductive base for establishing the category. The available information is represented by the two areas marked "known." A first problem for the N/I evaluator is filling in the remainder of the area. (The problem of fixing the boundary will be discussed later.) There appear to be three major strategies which he can pursue:

1. *Extension.* The inquirer begins with a known item or items of information and builds on them. He uses these items as bases for other questions or as guides in his examination of documents. Amoeba-like, he inches his way from the known to the unknown.

2. *Bridging.* The inquirer begins with several known, but apparently disconnected, items of information. The term "disconnected" simply means that their relationships are not understood. That there *are* relationships is a premise of high probability because the items have been placed into the same category. The evaluator now uses these two points of reference for further inquiry in an effort to identify the connections and understand them.

3. *Surfacing.* As the inquirer becomes more and more familiar with the area, he becomes able to propose new information that ought to be found in the field and then to verify its existence. This process of surfacing is similar to the familiar process of hypothesis formation, or to the process of suggesting new categories, once a subset of known categories has been identified, because the logic of the situation "demands" them.

Obviously, these processes are more art than science. But it certainly will not do to reject them simply because they cannot be scientifically explicated. All science depends heavily on the "arty" formulations of its practitioners. Depriving scientists of the right to use intuition would render scientific pursuit sterile. Scientific hypotheses are, after all, nothing more than the reconstructed logic of hunches, the logic-in-use.

Inclusion/Exclusion Criteria. The evaluator pursuing any strategy for "fleshing out" an area inevitably confronts decisions about which items of information to include and which to exclude. Listed below are several guidelines that seem appropriate. The first three relate directly to the strategies previously discussed; the other four are additional criteria that should prove useful:

1. Include any information item that extends the area of information. This automatic inclusion is tempered only by boundary setting conditions (explicated below).

2. Include any information item that relates or bridges several already existing items.

3. Include any information that identifies or brings to the surface new elements of importance.

4. Add any information which reinforces existing information, but reject it if the reinforcement begins to approach mere redundancy.

5. Add any information that tends to explain other information already available.

6. Add any information which exemplifies either the nature of the category or important evidence within the category.

7. Add any information which tends to refute existing information.

In general, items of information which do not fall within these limits should be excluded as redundant, irrelevant, or immaterial.

Closure. The final divergence question that faces the naturalistic evaluator is when to stop collecting information, i.e., a determination of when he has reached the outer boundary depicted in Figure 3. There seem to be four criteria to inform the *stop* decision:

1. *Exhaustion of sources.* When there are no new situations to observe, subjects to interview, processes to describe, or documents to analyze, etc., termination is patently indicated. But this occurrence is rare. Typically, there are more sources than can effectively be tapped, at least within the limited resources available to the evaluator. But even if the number of sources is limited, it is always possible, and probably desirable, to recycle, for example, in order to move through successive waves of discovery and verification, to continuously triangulate sources, to pursue strategies of extending, bridging, and surfacing described above which require recurrent contact with informants, and so on.

2. *Saturation.* When successive examination of sources (recurrent or new) tends to yield redundancy without useful reinforcement or produces seriously diminishing returns (i.e., produces small additional yield for substantial additional effort expended) it is probably time to terminate. Saturation may occur without the emergence of a sense of integration or completeness because the number of accessible or willing informants is small, because of a conspiracy of silence on the part of respondents, or because some crucial elements have failed to surface even though apparently logically required. But, except for the third of these conditions, there is little that the evaluator can do to re-open founts of information; the third case will test his wit and cleverness—attributes which have been known to fail even those who are liberally blessed with them.

3. *Emergence of regularities.* When the area "feels" integrated, as best attested by a sense of regularity in the available information, it is probably best to stop. But several caveats should be noted. Regularities may occur early even though some crucial items of information have not yet emerged. Moreover, regularities exist at different levels. A simplistic interpretation of information may appear to be highly regular but may be grossly in error at some other, more sophisticated level. Eisner's concept of connoisseurship is relevant here: some persons are more sophisticated than others. For these reasons it is probably best to invoke the criterion of regularity only in conjunction with one or more of the other criteria.

4. *Overextension.* Obviously, it is possible to collect information which goes beyond the boundaries at which it would be wise to stop if one only

knew that the boundaries existed. How can one tell whether one has moved beyond the pale of useful information, and so make a proper adjustment?

Barker and Wright (1955) have introduced the notion of "synomorphism" between the pattern of behavior being investigated and the milieu in which it is found. The milieu is said to be "circumjacent" to the behavior. If an inquiry is extended too far, it will be difficult to maintain a feeling of synomorphism between behavior and milieu, and the milieu will appear distant rather than circumjacent to the behavior. That is, the arena of investigation will appear far removed from the issue or concern which prompted it. When this is felt in the data it is a signal that the inquiry has been overextended.

None of these criteria is absolute, nor is even the combination of the four an infallible signal for closure. Nevertheless, the naturalistic evaluator can get reasonable guidance for a termination decision by attending to them. Continued inquiry in the face of exhausted resources, information saturation, data regularities, and overextension is clearly not a wise course of action.

PROBLEMS OF AUTHENTICITY

Problems of authenticity relate to the establishment of bases for trust in the outcomes of naturalistic inquiry. How might the naturalistic inquirer argue to persuade a methodologically sophisticated peer about the authenticity of his information, the categories he derives, and the interpretations he draws?

So far as *categories* are concerned in evaluation, questions of authenticity have probably been adequately dealt with in the previous discussion of *convergence problems.* This section deals with the authenticity of a naturalistic inquiry *as a whole* and specifically, any evaluation study which has been pursued along naturalistic lines. The subject of this discussion will be the criteria of *scientific* adequacy, not the more conventional criteria of utility that are often applied to evaluation studies such as timeliness, economy, tailored reporting, and the like. These have been well discussed by other authors.[35]

The classic criteria of authenticity applied to research are validity, reliability, and objectivity. As a matter of fact, all three of these concepts are currently undergoing an overhaul. Witness, for example, the redefinitions of validity and reliability that are being made in relation to criterion-referenced testing, and Scriven's brilliant restatement of the nature of objectivity (a perspective that will be reviewed below). Nevertheless, these concepts, slippery as they may be, are applicable to naturalistic inquiry, as in any inquiry,

[35]See for example Stufflebeam (1974), the Joint Committee on Standard for Educational Evaluation (1977), and Worthen (1977).

in the sense that they raise perfectly reasonable questions. The position to be taken here is that these concepts, while relevant, are in some need of reinterpretation in order to be fully applicable to the N/I situation.

In order to avoid confusion with the conventional connotations and denotations of these terms, new terms will be introduced: *intrinsic adequacy* in lieu of *internal validity*, *extrinsic adequacy* in lieu of *external validity* or *generalizability*, *replicability* in lieu of *reliability*, and *impartiality* in lieu of *objectivity* (see footnote p.44).

Intrinsic Adequacy (Internal Validity)

The degree of isomorphism that exists between the study data and the phenomena to which they relate is an important factor in establishing the intrinsic adequacy of a naturalistic inquiry. There appear to be a number of strategies for pursuing or assessing isomorphism.

Erecting safeguards against potentially invalidating factors. An analysis by Speizman (n.d.), based largely on materials included in a collection of essays edited by McCall and Simmons (1969), suggests the following invalidating factors may be present in an N/I inquiry:

1. *Distortions resulting from the researcher's presence at the research site.* Such distortions include the induction of reactive responses in subjects and the failure to provide sufficient opportunity for the researcher's preconceptions to "be thoroughly challenged by the data he collects." Speizman suggests that close monitoring of responses and a prolonged engagement at the research site are sufficient to overcome these effects.

2. *Distortions resulting from the field worker's involvement with his subjects.* Failure to establish a minimum level of rapport, on the one hand, or "over-rapport," on the other, can militate against a successful field study. Speizman counsels against too much involvement and proposes that the field worker constantly check himself to be sure that he is maintaining his objectivity.

3. *Distortions resulting from bias on the part of the field worker or his subjects.* Such distortions may arise from wrong first impressions, slavish adherence to hypotheses formed earlier, or role/status differentials. Subjects may introduce distortions for similar reasons, or simply out of a desire to be as "helpful" as possible. Speizman implies that special efforts and sensitivity on the part of the investigator will help to offset such tendencies.

4. *Distortions resulting from the manner in which data-gathering techniques are employed.* Speizman suggests that careful recoding of data, continual scrutiny of data for internal and external consistency, cross-checking of inferences with selected interview material, and continual assessment of subject credibility are sufficient to overcome most of these distortions.

While it is clear that the tactics that Speizman proposes would not finally establish the intrinsic adequacy of study findings to an outside observer, keeping these caveats in mind during the course of a study will undoubtedly be helpful in assuring the overall adequacy of the inquiry when it is put to a more definitive test.

Establishing the degree of structural corroboration. As Eisner (in press) suggests,

> Structural corroboration is a process of gathering data or information and using it to establish links that eventually create a whole that is supported by the bits of evidence that constitute the whole. Evidence is structurally corroborative when pieces of evidence validate each other. Take as an example of structural corroboration the work of Inspector Poirot in Agatha Christie's *Murder on the Orient Express.* What the inspector found himself with was a dead man, murdered on a train, and the problem of solving the puzzle of who murdered him. Little by little Poirot succeeds in putting the pieces together so that there are no contradictions, the pieces support one another and the problem of who killed the man is solved. In the case of the Orient Express it was everyone on the train with whom the murdered man once had contact. The final scene in the book finds Poirot presenting his case to the murderers in the lounge car of the train as it speeds toward Istanbul. Poirot's brilliant use of evidence works, his conclusions hold up.

Similarly, House (1977) points out that

> Validity is provided by cross-checking different data sources and by testing perceptions against those of participants. Issues and questions arise from the people and situations being studied rather than from the investigator's perceptions. Concepts and indicators 'derive from the subject's world of meaning and action.' In constructing explanations, the naturalist looks for convergence of his data sources and develops sequential, phase-like explanations that assume no event has single causes. Working backwards from an important event is a common procedure. Introspection is a common source of data.

There appear to be two useful techniques for establishing structural corroboration. Note that both techniques are used *during* the course of the inquiry and indeed, help to shape later stages of the inquiry by building on earlier stages. The techniques are:

1. *Triangulation.* Triangulation depends upon exposing a proposition (e.g., existence of an issue or concern; validity of some alleged fact; assertion of an informant) to possibly countervailing facts or assertions. Denzin (1971) observes:

> Triangulation forces the observer to combine multiple data sources, research methods, and theoretical schemes in the inspection and analysis of behavioral specimens. It forces him to *situationally* check the validity of his causal propositions. . . . It forces him to *temporarily* specify the

> character of his hypothesis. . . . It directs the observer to compare his subject's theories of behavior with his emerging theoretical scheme. . . . The naturalist must have an intimate familiarity with all his data sources so he can judge which ones to discount, which ones to treat as negative cases, which ones to build into his representative cases. (p. 177)

Webb *et al.* (1966) conclude that triangulation, though difficult, is very much worth doing, because it makes the data and findings believable:

> Once a proposition has been confirmed by two or more measurement processes, the uncertainty of its interpretation is greatly reduced. The most persuasive evidence comes through a triangulation of measurement processes. If a proposition can survive the onslaught of a series of imperfect measures, with all their irrelevant error, confidence should be placed in it. (p. 3)

The phrase, "imperfect measures, with all their irrelevant error," is especially illuminating. Obviously, the naturalistic investigator cannot put very much confidence into particular observations or deductions. Each will contain its modicum of error, perhaps sufficient to cause suspension of belief if it were the only one. But when a series of bits of evidence all tend in some direction, that direction assumes far greater believability. As statistical means are more stable than single scores, so triangulated conclusions are more stable than any of the vantage points from which they were triangulated.

2. *Cross-examination.* In the law, a major purpose of cross-examination is to establish that there are alternative inferences which may be drawn from any set of "facts." It is this aspect of cross-examination that gives it unique utility for determining the degree of structural corroboration. If intensive cross-examination cannot shake the web of facts and inferences that have been drawn, that web has structural integrity.

Tymitz and Wolf (1977), in discussing cross-examination as a feature of the legal model of evaluation, propose the following specific objectives for it:

> 1. To establish the witness' frame of reference or bias so that the values and/or assumptions underlying her/his testimony will be clear. This will enable the panel to understand the testimony's context, its subtleties and nuances and perhaps even to understand why the witness believes the way s/he does.
>
> 2. To point out the flaws, alternative consequences or trade-offs for any recommendations the witness offers.
>
> 3. To clarify, extend, or modify facts, opinions, or beliefs expressed in direct examination.
>
> 4. To help the panel understand that plausible inferences exist other than the one(s) established by direct examination for any piece of evidence.
>
> 5. To seek justification for certain judgments. (pp. 59-60)

If, in this statement of purposes, the terms "panel" and "witness" are replaced with "investigator" and "subject," a description of the use of the

cross-examination technique appropriate for establishing structural corroboration emerges.

Establishing adequacy through persistent observation. If one way to establish the adequacy of a "fact" or inference is through the use of repeated perspectives, as described above, another way is through the use of repeated observations from a single perspective. In discussing the connoisseurship model of evaluation, Eisner (1975) notes:

> When one deals with works of visual art and works of literature there exists a certain stability in the material studied. But what do we do with things and events that change over time; classrooms, for instance? How can something as fluid as a classroom be critically described and how can such descriptions be tested for their referential adequacy? . . . The classroom being studied needs to be visited with sufficient persistency to enable the critic to locate its pervasive qualities; those qualities through which aspects of its life can be characterized. Classrooms or schools are not so fugitive that their pervasive qualities change on a daily basis. What is enduring in a classroom is more likely to be educationally significant than what is evanescent. These enduring or pervasive qualities can become objects of critical attention. (p. 18)[36]

In another context Eisner (*in press*) makes the point that

> One of the reasons why it is important . . . to have extended contact with an educational situation is to be able to recognize events or characteristics that are atypical. One needs sufficient time in a situation to know which qualities characterize it and which do not.

Thus, internal adequacy is, to some extent, a function of the amount of time and effort which the naturalistic inquirer invests in repeated and continuous observation. Not only will the investigator be able to differentiate typical from atypical situations or identify the enduring or pervasive qualities which characterize a situation, but he will also know when to give credence to the occasional aberrant or apparently idiosyncratic observation which nevertheless carries great insight and meaning. Thus, persistent observation and extended contacts are some of the hallmarks of the internally adequate evaluation.

Establishing credibility of findings. Since so much of naturalistic inquiry depends upon the perceptions of informants, it is essential that they find the data and inferences of a naturalistic study credible and persuasive. Eisner (*in press*) has called this approach "multiplicative corroboration–the use of . . . peers to pass judgment on what has been structurally corroborated."

---

[36] Eisner's term "referential adequacy" encompasses not only persistent observation but also the use of specially prepared criterion materials. See p. 66.

Assurance of credibility of the final result of a naturalistic inquiry is probably best obtained through frequent and thorough interaction with informants as the information develops. In this fashion information with limited credibility can be identified early and either eliminated or buttressed. Of course, one may object that such a process exposes the investigator to untoward influences. While this possibility is undoubtedly real, the investigator can guard against it by instituting the safeguards suggested on pp. 62 ff.

Donald Campbell (*in press*) has proposed what he calls "participant evaluation," a concept that seems closely related to the use of participants for providing credibility checks:

> Participants . . . will usually have a better observational position than will anthropologists or other outside observers of a new program. They usually have experienced the preprogram conditions from the same viewing point as they have the special program. Their experience of the program will have been more relevant, direct and valid, less vicarious. Collectively, their greater numerosity will average out observer idiosyncracies that might dominate the report of any one ethnographer.
>
> While participants are asked to generate a lot of data in program evaluation, rarely are they directly asked to evaluate the program, to judge its adequacy, to advise on its continuance, discontinuance, dissemination, or modification. Rather than evaluating programs, participants are usually asked about themselves and their own adequacy. We are thus wasting a lot of well-founded opinions.

Although Campbell addresses program evaluation in the above passage and suggests the use of participant judgments as part of the evaluation itself rather than as an evaluation of the evaluation, it seems clear that his observations would also hold in the latter case.

It is likely that the criterion of respondent credibility is the single most important judgment that can be brought to bear on a naturalistic inquiry. Without it one can have no sense that the findings and inferences have any reality, particularly since so much depends upon the perceptions of people. With it, except in the case of a general conspiracy to mislead the investigator, one can be reasonably sure that the findings do reflect the insights and judgments of a large group of people coming from different perspectives.

Establishing congruence with specially prepared referential adequacy materials. Eisner (1975), as part of his treatment of referential adequacy, interestingly suggests that it may be possible to compare an evaluation report with certain specially prepared materials that will allow an external observer to make his own judgment about the instrinsic adequacy of the report. In relation to classroom observations, he indicates that

> the availability of videotape recordings and cinematography now make it possible to capture and hold episodes of classroom life that can be

> critically described. Such videotaped episodes can then be compared with the criticism created and its referential adequacy determined. In addition, playback features of videotape make it possible to scrutinize expression, tempo, explanation, and movement in ways that live situations will not permit. Disputes about the adequacy of criticism can be resolved, at least in principle, by re-examining particular segments of the tape. The technology now available lends itself exceedingly well to the work to be done. (pp. 18-19)

Eisner's proposal is not dissimilar to the call from Barker (Willems & Raush, 1969, pp.37ff.) and his associates for data archives. Not only would such archives provide original study material, but they could also be used to make the kinds of comparisons Eisner suggests with videotapes and films.

It is possible that special techniques particularly suited to this approach might be worked out. The materials that both Eisner and Barker describe are largely fortuitous, collected during the investigation itself and primarily as original data, not for cross-validational purposes. There seems to be no reason, however, why sampling designs could not be set up which would furnish materials collected particularly for this latter function. They would be essentially parallel materials (not too different from the parallel test forms frequently used to establish test reliability) against which the adequacy of the fuller, more complete study materials could be tested.

Extrinsic Adequacy (External Validity)

For many evaluation purposes the question of extrinsic adequacy is meaningless. The teacher evaluating the extent to which his class has mastered certain concepts, the school system endeavoring to determine which of several available textbooks is most appropriate for its pupils, the developer deciding which of several presentational formats to use, all provide examples of particularistic questions from which generalizations are unlikely to be made. Indeed, it is probably the case that in most formative evaluations and in certain summative situations in which the interest of the evaluator is focussed on a particular time or situation (e.g., a particular class), questions of generalizability are irrelevant.

Nevertheless there are many situations in which generalizability is an issue, or at least might be considered an issue. What might the naturalistic evaluator do under such circumstances? The answers to that question range from "not much" to "a good deal."

Before turning to that matter, however, one observation is in order. It is a truism that external validity cannot exist unless there is a reasonable level of internal validity. Clearly, there is no point in asking whether meaningless information has any general application. In that sense, internal validity reinforces external validity. But there is also a sense in which internal validity

can truncate or inhibit external validity. An overemphasis on *a priori* control of factors or conditions influencing the inquiry, which have been instituted precisely in the interest of achieving high internal validity, may seriously affect the external validity because the findings can then, at best, be said to be generalizable only to other, similarly controlled situations. Since these situations rarely exist in real life, generalizability can generally be questioned in the case of tightly controlled experiments, simply because they are conducted under laboratory conditions. Most of the methodologies reviewed in Chapter 2 have emerged, in part, because of reaction to this criticism, particularly those of Barker and other eco-psychologists.

Naturalistic inquiry and naturalistic evaluation seem exempt to this criticism because no attempt is made to institute *a priori* controls either on variables entering into the investigation or on output measures. There is full congruence or isomorphism between the situation being explored and the real world about which generalizations are to be made. Thus, the N/I inquirer might claim that his inquiry has built-in, or guaranteed, extrinsic adequacy.

Unfortunately, the problem is not that simple. Whether or not certain information is generalizable is a function of the degree to which the situation being investigated is a representative "slice of life," as well as a function of the isomorphism between the situation in which it was generated and the situation to which it is to be generalized. But, at least the N/I investigator *is* free of one problem that continuously plagues the experimentalist: the lack of isomorphism between laboratory and real world.

There are several positions that might be taken in regard to generalizability:

1. *Generalizability is a chimera; it is impossible to generalize in the scientific sense at all.* The person espousing this view argues that there are multiple realities in every situation. Different people with different backgrounds and purposes will see different things when they view a situation. The onion has many layers, and it is difficult for two people to come to any agreement about which layer is the one to view. Replication is impossible because things change so rapidly. A person viewing the same situation at different times will be forced to different conclusions. Situational interactions are always so complex that any observation can have meaning only for the actual situation in which it occurred. If the term "generalization" is to have any meaning at all, it must be for particular audiences. It is up to each audience to determine what, if anything, the information means, and to determine for itself its applicability. The principle burden for synthesis must always lie with the recipient of an evaluation report or research study; it cannot lie with the evaluator or the researcher.

Certainly each of these points has validity. But the position is extreme—it appears to be less a genuine position than a reaction to the over-emphasis

placed on generalizability by conventional methodologists. The pendulum can swing very hard when it swings. Counter-cyclical tendencies are useful, but over-reactions simply antagonize persons looking for serious and credible methodological alternatives.

2. *Generalizability continues to be important, and efforts should be made to meet normal scientific criteria that pertain to it.* This is a second view that might be taken. Denzin (1971), for example, indicates that "to the best of his ability the investigator must offer evidence on the degree to which his samples of behavior are representative of the class of joint acts he wishes to generalize to" (p.175). However, he quickly notes that the population to which he wishes to generalize may be unknown; ". . . as a consequence the naturalistic observer seldom can specify with precise detail the universe of interactive relationships to which he wishes to generalize" (p.175). Of course, that observation holds for most behavioral inquiries. It is certainly not unique to N/I. Denzin does suggest several approaches that can be taken, as a function of the investigator's focus.

a. *A representative situation.* The investigator locates himself in a "representative" situation and argues generalizability to all persons who pass through that situation.
b. *A time* (e.g., a particular season, month, holiday). The investigator argues that his findings hold for all persons in that time frame.
c. *Special populations* (e.g., schoolchildren, an incarcerated group). The investigator generalizes to a class of social organization (e.g., schools or prisons).
d. *A recurring encounter* (e.g., a police-juvenile interaction). The investigator generalizes to types of encounter, or to the population of which the persons involved in the encounter is a "sample."

While these are laudable approaches and perhaps should be utilized whenever opportunities present themselves, they nevertheless fall sufficiently short of the theoretical ideal so that the naturalistic investigator cannot definitively claim generalizability. The effect is analogous in power to the tactic of matching groups as a replacement for random selection and assignment. Some gains undoubtedly occur, but the basic question, how good a sample of the population is being dealt with, remains essentially unanswered.

3. *Generalizability is a fragile concept whose meaning is ambiguous and whose power is variable.* This third possible argument is essentially the one made on pp. 28ff. based largely on the assertions of Cronbach (1975) that generalizations decay; that in a short while after they are made, generalizations are more likely to be history than science; and that instead of making generalization the ruling consideration in inquiry, emphasis should be placed on careful description, with efforts to "describe and interpret the effect anew

in each locale, perhaps taking into account factors unique to that locale or series of events" (p.125). Generalizations in this context have the form of working hypotheses, not of conclusions.

Cronbach's advice seems particularly appropriate to naturalistic evaluators. Since their major concern is often not generalization anyway, a willingness to move away from rigorous criteria of generalizability seems a small loss. Moreover, the rigor of generalizability is in all events suspect. If deferred generalization is a hallmark of a good inquiry model as described by Cronbach, the evaluator whose activity often leads to judgments which may have major consequences ought to be more interested than most in deferring. An error on his part is not simply a theoretical mistake but a possible disaster in the making.

A proposed resolution. What can be concluded from these varying points of view? The concept of generalizability clearly is undergoing revision. It is not valued as much as formerly, and its meaning has changed somewhat. The evaluator should do what he can to establish the generalizability of his findings: insuring that he has good intrinsic adequacy; using the Denzin tactics for building credibility when possible; repeating observations in each locale, and doing his best to identify the interactive elements in the situation which change the meaning of the generalization from its last point of emergence. Often naturalistic inquiry can establish at least the "limiting cases" relevant to a given situation. But in the spirit of naturalistic inquiry he should regard each possible generalization only as a working hypothesis, to be tested again in the next encounter and again in the encounter after that. For the N/I evaluator, premature closure is a cardinal sin, and tolerance of ambiguity a virtue.

Replicability (Reliability)

The issue of replicability is frequently invoked by critics of naturalistic inquiry. Willems (Willems & Raush, 1969) observes that:

> Naturalistic research is often criticized on the grounds that it is not replicable, and it often seems to be assumed that behavior represents a class of such unstable and complicated phenomena that unless it is constrained by experimental controls, it is not amenable to scientific study. (p. 57)

Replicability, like intrinsic adequacy, is often a non-issue for the naturalistic inquirer. The implications of a multiple reality on generalizability have been touched on in several contexts. Similar arguments apply in the case of reliability. Moreover, the naturalistic inquirer may often be more interested in differences than in similarities. The odd case may intrigue him more than the

recurrent regularity. Finally, it may be argued that reliability, as an issue, need not be faced directly. Since it is impossible to have intrinsic adequacy without reliability, a demonstration of intrinsic adequacy amounts to a simultaneous demonstration of reliability. Hence, if the evaluator places his effort in shoring up validity, reliability will follow.

Whatever the merits of such arguments, they do not answer the question that deserves an answer: what can be done to demonstrate replicability if a naturalistic inquirer chooses to do so?

Replicability of studies is an empirical matter. Evaluations especially, but often other naturalistic studies as well, are singular: they are commissioned and carried out to inform a decision, to respond to a criticism or concern, to determine the effectiveness of a method. Thus, the likelihood that such a study will normally be replicated is small, and almost no instances can be found in which replications have been undertaken. Are there empirical techniques which might be applicable? There seem to be three:

1. *The Audit.* The point was made, with respect to categorical sets (p.56), that one criterion of the necessity and sufficiency of such a set is its reproducibility by another competent judge. While it is too much to expect that sets of categories made up by two independent judges from the same basic data would coincide (for the reasons of multiple reality), a second judge should be able to verify that: (a) the categories devised by a first judge make sense in view of the data from which he worked, and (b) the data have been appropriately arranged into a category system. The second judge *audits* the work of the first much like an examiner audits the work of an accountant.

Similarly, a study may be audited by a second investigator or team of investigators. The second team could not be expected to reproduce the study without knowing the decisions made by the original investigators at every step of the process. They could, however, review each decision and the consequent actions based upon the first study team's knowledge at the point of that decision, verifying that substantively and methodologically sound options were chosen. Such a review or audit would give substantial assurance of the reproducibility of any inquiry.

2. *Overlap methods.* Imperfect inquiry techniques such as quasi-experimental designs can achieve enormous power when used in tandem, in such way that the imperfections of one are cancelled by the strong points of a second, complementary technique.[37] While such nested use of techniques is ordinarily recommended for the sake of validity, it is obvious that it can also be used to bolster the case for reliability. Comparable results from two or more different approaches strengthen the reliability claims of each of the individual approaches.

[37]This idea is frequently advanced in Webb *et al.* (1963). See also Campbell and Stanley (1963).

Reference was made earlier to Campbell's call for participant evaluation (p.66). In connection with his remarks on that point, Campbell (*in press*) reviews some recent programs or proposals in which participant judgments might have utility: the Head Start program, the negative income tax experiments, the judicial hearings proposed by Wolf and Levine, and his own studies of social service delivery systems. He then asserts:

> We need to try out a variety of such procedures. . . . these would represent methodologically independent cross-validations of the quantitative results. They would have the chance of discovering program effects on topics not anticipated in the formal measurement devices. *They would be likely to confirm the major findings on shared dimensions.* (emphasis added) (p. 28)

While Campbell's remarks are directed at the issue of validity (and the validity of quantitative measures at that), it is plain that he shares the conviction that overlapping methods also have the capability of confirming (repeating) major findings. Thus, in the interest of demonstrating replicability, the naturalistic evaluator should forage for information by using several methods simultaneously whenever possible. In a sense, this approach is simply a variant on triangulation, a procedure with which the N/I investigator is quite familiar.

3. *Stepwise replication.* It is crucial that some effort be made in each evaluation, whether based on single or multiple methods, to build replications into the several stages or steps of the work. Probably the most straightforward means for doing so is to divide both the information sources (interview subjects, observational situations, documents, and so on) and the study team (if it consists of more than one person) into two random halves. Each team then undertakes an independent study.

A major problem with this approach is that N/I designs are emergent, not preordinate, so that it is not possible to keep the two teams separated for the entire period. To obviate this serious problem it is necessary for the two teams to communicate with each other at the conclusion of each step, and probably at the conclusion of each work day. At early stages in which the intent is to identify issues and concerns, the teams may, after a period of independent work, share the issues and concerns they have unearthed to that point. The purpose of such sharing is partly to run the necessary replicability cross-checks and partly to set a common stage for the next step—to verify the issues and problems that *both* teams have identified. In the replication cross-check, it is up to each team to convince the other that issues and concerns identified by that team, but not by the other, are valid. In that sense, this team cross-checking is very similar to the auditing process described above, differing from it mainly in that the check is run during the work rather than at its end, and by team members rather than outside

auditors. It seems apparent that the reliability of a study would be well established if it could be shown that, at every crucial stage, the independent teams corroborated one another's work or could, in the auditing mode, be persuaded that one another's work had adequacy.

Neutrality (Objectivity)

The issue of objectivity is probably the most thorny one that can be raised with respect to naturalistic inquiry. For how can any inquiry be objective if it simply "emerges"; if it has no careful controls laid down *a priori*; if the observations to be made or the data to be recorded are not carefully specified in advance; if, on the admission of its practitioners, there exist multiple realities capable of being plumbed to different depths at different times and by different investigators?

On close examination, the difficulty noted here stems less from the innate characteristics of naturalistic inquiry than from the meaning which is ascribed to the term "objectivity." Scriven (1972) has pointed to a "fundamental confusion" in the use of the term which he describes as follows:

> The terms "objective" and "subjective" are always held to be contrasting, but they are widely used to refer to two quite different contrasts, which I shall refer to as the *quantitative* and the *qualitative* senses. In the first of these contrasts, "subjective" refers to what concerns or occurs to the *individual* subject and his experiences, qualities, and dispositions, while "objective" refers to what a *number* of subjects or judges experience—in short, to phenomena in the public domain. The difference is simply the *number* of people to whom reference is made, hence the term "quantitative." In the second of the two uses, there is a reference to the *quality* of the testimony or the report or the (putative) evidence, and so I call this the "qualitative" sense. Here, "subjective" means unreliable, biased or probably biased, a matter of opinion, and "objective" means reliable, factual, confirmable or confirmed, and so forth. Now it would certainly be delightful if these two senses coincided, so that all reports of personal experience, for example, were less reliable than all reports of events witnessed by a large number of people. But as one thinks of the reliability of reports about felt pain or perceived size, on the one hand, and reports about the achievements of stage magicians and mentalists, on the other, one would not find this coincidence impressive. (pp. 95-96)

The essentials of Scriven's argument are depicted in Figure 4. The two senses in which the terms "subjective" and "objective" can be used define the rows of the figure. The key differentiations which Scriven makes are given in the cells. The quantitative difference between subjective and objective is dependent solely on the number of subjects or judges involved—one in the subjective cell and more than one in the objective cell. The qualitative sense of subjective and objective depends upon the relative bias and confirmability of the report—the subjective cell reflecting greater bias and less confirmation.

The crucial points that Scriven makes involve the relationships among the cells of the figure, which have, for convenience, been labeled *A* through *D*. Several of these relationships are worthy of specific note:

1. A ≠ C nor does B ≠ D, *necessarily*. Although it is often assumed that individual concerns or thoughts are biased and unreliable, and that group experiences are reliable and confirmable (at least in the sense of inter-subject agreement), this assumption is not necessarily true. Depending upon the qualities, training, and experience of the single observer, what he reports might well be more reliable and confirmable than the reports of a group.

2. A *can* = D and B *can* = C. Judgments of a group can be biased and unreliable while judgments of individuals can be factual and confirmable.

3. A *can* = B, but C *cannot* = D. It is of course quite possible that an individual's experiences can be the same as the experiences of a group of which he is a member. Indeed, if this were not possible, the whole concept of inter-subjective agreement would be meaningless. In the *quantitative* sense, then, subjectivity and objectivity can be identical (although it may be impossible to supply a test to demonstrate that conclusively). In the *qualitative* sense, subjectivity and objectivity can *never* be identical; they are polar opposites.

In any inquiry, the objectivity of the data is of critical concern. The data should be reliable, factual, and confirmable. There seems to be no intrinsic

	Subjective	Objective
Quantitative	What an individual experiences (N = 1) A	What a number of individuals experience (inter-subjectivity) (N > 1) B
Qualitative	Unreliable, biased, matter of opinion C	Reliable, factual confirmable or confirmed D

Figure 4: Representation of Quantitative and Qualitative Aspects of Objectivity/Subjectivity

Based on Scriven (1972)

reason why the methods of a properly trained naturalistic inquirer should be any more doubtful a source of such data than the methods of an investigator using a more quantitative approach. After all, data gained from quantitative sources may also be biased. Consider for example, the cultural bias said to exist in so many "objective" tests. The issue is *not* the *intrinsic* objectivity (in the qualitative sense) of the method used to generate information, but the *confirmability* of the information once it is obtained. In that regard, the methods for establishing intrinsic adequacy discussed earlier seem appropriate: triangulation, cross-examination, persistent observation, credibility checks with informants and audiences, and testing of congruence with special referential adequacy materials.

Threats to neutrality. But, while all of the preceding may argue for a different view of objectivity, here called "neutrality" to distinguish it from its historical precursor, the naturalistic evaluator cannot ignore the establishment of some safeguards. There are at least four threats to an inquirer's objectivity:[38]

1. *Unconscious bias.* The evaluator may bring into the situation certain unconscious biases which cause him to interpret observations in a particular way. For example, unconscious sexism may lead him to discount input from female respondents or, conversely, to weight input from male respondents more heavily. Bias is perhaps the most pernicious of the threats to objectivity. Being unconscious, it is most difficult to detect and most difficult for the evaluator to acknowledge if it should happen to be called to his attention.

2. *Conscious prejudice.* Conscious prejudice, like unconscious bias, causes misinterpretations. Precisely because it is conscious, such prejudice is more easily identified, but unfortunately may be more difficult to remedy. Often unconscious bias can be corrected simply by calling the evaluator's attention to it. People with confirmed prejudices tend to retrench when challenged. It is of course reprehensible for an evaluator to behave in prejudicial ways, but in such cases one must assume the presence of psychologically compelling reasons for the behavior. Thus the problem is not easily surmounted.

3. *Incompetence.* Evaluators are just as subject to incompetence as any other group of professionals. Incompetence can arise from ignorance or from irresponsibility; both result in shoddy practice.

4. *Gullibility.* Almost everyone involved with or affected by an evaluation has something at stake in its outcome. Evaluators should expect both conscious and unconscious attempts to influence their data collection and their interpretations. The evaluator who takes testimony at its face value is simply naive.

---

[38]I am indebted to Fred Ellett for calling my attention to these threats to neutrality.

5. *Corruptibility.* Evaluators can be and are corruptible, for both good and bad reasons. An evaluator, for example, may enter into a conspiracy with a client to produce the kind of judgments that the client wants. Or the evaluator may choose to shade the data with good intentions, e.g., to protect the morale of a practitioner group, to avoid upsetting delicate political balances, to avoid closeout of a program which, while perhaps not accomplishing the objectives it set for itself, nevertheless produced a number of positive side-effects. Or the evaluator may be more negative in his interpretations than he needs to be, to justify a decision already made on other grounds, to provide leverage for soliciting funds to support an effort to attack some problem which he or his client are determined to attack, or simply to demonstrate his own tough-mindedness, a situation frequently encountered when third-party evaluators are called in.

Safeguards against these threats are difficult to establish. In part the methodology provides protective strategies. The evaluator who triangulates his data sources, cross-checks his testimony, is persistent in his observations and, in short, follows the normal procedures of naturalistic inquiry, has done all he can in the course of the investigation. *Ex post facto* audits may provide some assurance to outside audiences that objectivity criteria were reasonably met. Beyond that, the best guarantees of an objective evaluation that can be provided to a client or an evaluation audience are that the evaluator whom they have hired is adequately trained, experienced, and reputable. The profession also has a responsibility to provide a strong code of ethics and to police itself.

Two other concepts are relevant to the notion of neutrality, *openness* and *fairness.*

Openness. Openness is difficult to achieve. Obviously, there is no procedural tactic that will insure openness nor is there any test that can be applied after the fact to permit the judgment of whether openness was actually attained. Nevertheless, the N/I evaluator, in working with informants (or indeed, with other data sources) should avoid the following kinds of erroneous assumptions:

1. *Experiential analogies.* The evaluator may erroneously assume that the experience of his informant is like his own and that there exists a common background within which to interpret phenomena. An informant who expresses a concern about the introduction of new math may be reflecting his own experiences in the life adjustment schooling of the 40's and 50's. The nature of his objection can be seriously misunderstood by the evaluator who interprets his remark in the context of his own experience in the national curricula that characterized the 60's.

2. *Psychological analogies.* The evaluator may assume that the psychological characteristics and attributes of his informant are similar to his own. The anxiety engendered by an evaluation can hardly be fathomed by the outgoing, aggressive evaluator. The desire of the informant to say and do things that the evaluator will find useful may be mistakenly interpreted as honest data.

3. *Conceptual analogies.* The evaluator may believe that the informant has the same understandings as he. The informant's comments about organizational conflicts, based upon a one-time reading of a *Reader's Digest* article, do not have the same meaning as the evaluator's own well-articulated concepts based upon a thorough knowledge of organizational literature. Conversely, and probably more importantly, the informant is often better-versed in the area being evaluated than is the evaluator. The evaluator must be careful not to assume that he understands all the subtle distinctions that exist in the mind of the informant.

Fairness. The concept of "fairness" as an important aspect of evaluation has been posed most articulately by Ernest House (1976). In relation to objectivity, he raises some interesting questions:

> Is it really sufficient to say that an evaluator is objective? If objectivity is taken in the commonly-used sense of employing an externalized, specifiable procedure which produces replicable results then it is certainly an insufficient criterion for an evaluation. The administration of standardized achievement tests is a totally externalized, specifiable procedure which produces replicable results. At the same time such tests are thought to be highly biased in many ways, particularly toward minority groups. In this sense, one has an objective but biased instrument. In fact, one can produce an instrument in which the bias is in the other direction. (To further confound matters, if racial discrimination is the intent of such an instrument, one could have an objective, valid instrument for that purpose.)
>
> An evaluation must be free from distortion and bias (qualitatively objective) and being externalized, specifiable, and replicable does not sufficiently address possible biases. (p. 75)

House proposes that evaluation be carried out in accordance with two principles of justice, which he has, in turn, borrowed from Rawls (1971):

First Principle

> Each person is to have an equal right to the most extensive system of equal basic liberties compatible with a similar system of liberty for all.

Second Principle

> Social and economic inequalities are to be arranged so that they are both:
>
> (a) to the benefit of the least advantaged, consistent with the just savings principle, and

(b) attached to offices and positions open to all under conditions of fair equality of opportunity.

How, asks House, would an evaluation which was administered under these principles differ from a more conventional evaluation?

> According to the first principle of justice, the basic liberties of the parents, teachers, and students, including self-esteem, could not be infringed upon. Any method which raised test scores at the expense of self-esteem would be judged as wrong—for any one of the concerned groups. If the self-esteem of one group depended on detracting from the self-esteem of another, that is simply not right.
>
> By the second principle, social and economic inequalities must benefit the least advantaged in the long run. The educationally least advantaged within most settings are the children first and the teachers second. The evaluator should strive to present their views and perspectives. In addition, there are the least advantaged in the whole of society. Insofar as the educational gains are translatable into social and economic advantages to a particular group at the expense of others, the program must be justified on the contention that the inequalities are to the benefit of the least advantaged.
>
> At least once within the priorities of the two principles as applied, the program should be regarded from the perspective of the educationally least advantaged and the socially least advantaged. In conflicting situations over inequalities, the interests of the least advantaged are to take precedence—consistent with the other principles. (p. 97-98)

Thus, the concept of fairness seems to involve responding to concerns, particularly of disadvantaged groups. The evaluator has an obligation to uncover those concerns, even when there is no one coming forth to express them. The evaluator who makes a determined effort to be fair is likely to produce a more unbiased and confirmable evaluation than one who simply allows information to emerge when and if there is an individual or group who brings it to his attention.

Chapter V:

SUMMARY

This monograph has made a statement about naturalistic inquiry in educational evaluation. The points covered can be briefly summarized:

1. Although evaluation has come to be widely viewed as the handmaiden of decision-making and social policy development, it has not been as fruitful as hoped or expected. While there are, no doubt, many reasons for this failure, one of them is the lack of a methodology uniquely suited to evaluation's needs. Naturalistic inquiry is proposed as one such alternative.

2. Naturalistic inquiry is definable by what the investigator *does*. Investigator actions can take two forms: imposition of *a priori* constraints on the antecedent conditions (control), and imposition of constraints on the outputs. The more an investigator does both, the more he is experimentally inclined; the more an investigator does neither, the more he is naturalistically inclined. Both experimental and naturalistic approaches (as well as others) can, thus, be conceptualized as occupying different areas within a common inquiry space defined by the two dimensions noted (see Figure 2).

3. Defining inquiry in this way provides no clues to why an investigator would choose one form of inquiry over another. In general, that decision is made by explicit or implicit reference to certain other dimensions which characterize inquiry (see Table 1). These dimensions include the philosophical base (logical positivism vs. phenomenology), inquiry paradigm (experimental physics vs. ethnography or investigative journalism), purpose (verification vs. discovery), stance (reductionist vs. expansionist), framework/design (preordinate vs. emergent), style (intervention vs. selection), reality manifold (singular vs. multiple), value structure (singular vs. pluralistic), setting (laboratory vs. nature), context (unrelated vs. relevant), conditions (controlled vs. invited interference), treatment (stable vs. variable), scope (molecular vs. molar), and methods (objective–intersubjective agreement vs. objective–factual/confirmable).

4. The need for naturalistic inquiry may be argued on the basis of the following purposes: to enlarge the arsenal of investigative strategies available for dealing with emergent questions of interest; to provide an acceptable basis for studying process, to provide an alternative where it is impossible to meet the technical assumptions of the experimental approach in the real world, to better assess the implications of treatment-situation interaction, to redress the

balance between reconstructed logic and logic-in-use, to avoid the implicit shaping of possible outcomes, to optimize generalizability, and to meet certain practical criteria defined as fitting, working, and communicating.

5. Naturalistic inquiry has a certain special utility for the practice of evaluation because of a strong, current scholarly and theoretical interest in moving away from more conventional paradigms; a strong movement among evaluation practitioners away from conventional methodologies toward more naturalistic approaches; and the emergence of a number of evaluation models which seem especially congenial to the use of naturalistic inquiry as a supporting tool or technique (including the responsive model, the judicial model, the transactional model, the connoisseurship model, and the illumination model).

6. Naturalistic methodology, like all methodology, represents an article of faith on the part of its practitioners and users. There is no compelling way to truth; all truth depends, in part, on persuasion. If naturalistic evaluation is to be credible, its practitioners must recognize the standard methodological problems and deal with them in a way that is convincing to evaluation audiences and critics.

7. Naturalistic inquiry does confront certain standard problems, defined here as boundary problems, focussing problems, and problems of authenticity.

8. Boundary problems relate to the means by which the naturalistic inquirer sets limits to his inquiry. These limits emerge from the nature of the problem (or the mandate set by a client) which the investigator chooses to study. A problem format may be defined which limits the range of inquiry activity to verification, causation, contravention, and consequence.

9. Focussing problems are defined as of two varieties: convergence problems, relating to the establishment of sets of categories within which data can be assimilated and understood; and divergence problems, relating to the "fleshing out" of those categories with sufficient information for completeness. It is proposed that the issues and concerns in the minds of relevant audiences be taken as the basis for establishing categories; and guidelines for unitizing and categorizing are offered. Strategies are proposed for the "fleshing out" activity (extension, bridging, and surfacing); some criteria for inclusion/exclusion are proposed; and some guidelines for establishing closure are provided.

10. Problems of authenticity are the most sticky ones confronting the naturalistic inquirer. In dealing with problems of intrinsic adequacy (internal validity), the naturalistic inquirer is advised to establish certain safeguards against potentially invalidating factors; to establish structural corroboration through such techniques as triangulation and cross-examination; to engage in

persistent observation; to run audience credibility checks, and to run checks against special referential adequacy materials. With respect to extrinsic adequacy (external validity), the naturalistic evaluator is advised (after a review of possible positions regarding generalizability) to adopt Cronbach's position on the softness of generalizability and to treat each possible generalization only as a working hypothesis to be tested and retested. With respect to replicability (reliability), the naturalistic inquirer is advised to avoid easy arguments that reduce replicability to a non-issue and, instead, to utilize several techniques that will help make the case for replicability: the audit, overlapping methods, and stepwise replication. Finally, with respect to neutrality (objectivity), Scriven's key distinctions between quantitative and qualitative definitions of objectivity are invoked. It is argued that conventional objections to naturalistic inquiry based on a quantitative concept of objectivity (inter-subject agreement) are inappropriate and that whether or not an N/I inquirer is biased depends a great deal on his training and experience. Bias, prejudice, incompetence, gullibility, and corruptibility are indicated as possible threats to neutrality. Additionally, some attention is paid to the concepts of openness and fairness as relevant, qualitative criteria.

Naturalistic methodology is by no means fully explicated. The title of this monograph, "*Toward* a Methodology of Naturalistic Inquiry in Educational Evaluation," was chosen with that *caveat* in mind. There is no pretence that what has been said here is definitive. But it does seem to the author that N/I is methodologically sufficiently well-grounded to warrant trial. The inevitable methodological questions can be answered well enough to permit reasonably sound operations in the field. The method is well suited to applications in evaluation, especially as that field is being defined in emergent models. It may be difficult to persuade some audiences (particularly those who insist on impact evaluation based on test scores) to accept the procedures suggested here in lieu of what they understand to be evaluation. Yet, it also seems that naturalistic inquiry offers a more congenial and responsive mode of evaluation than any other practiced today.

BIBLIOGRAPHY

Abt, C. C. *The evaluation of social programs.* Beverly Hills, CA: Sage Publications, 1977.

Argyris, C. Dangers in applying results from experimental social psychology. *American Psychologist*, 1975, *30.*

Barker, R. G. Explorations in ecological psychology. *American Psychologist*, 1965, *20.*

Barker, R. G. *Ecological psychology.* Stanford, CA: Stanford University Press, 1968.

Barker, R. G., & Wright, H. F. *Midwest and its children.* New York, NY: Harper and Row, 1955.

Becker, H. S. Problems of inference and proof in participant observation. *American Sociological Review*, 1958, *58.*

Bogdan, R., & Taylor, S. J. *Introduction to qualitative research methods.* New York, NY: John Wiley and Sons, 1975.

Brickell, H. M. Needed: Instruments as good as our eyes. *Journal of Career Education*, 1976, *2.*

Bronfenbrenner, U. The experimental ecology of education. *Educational Researcher*, 1976, *5.*

Bronfenbrenner, U. Toward an experimental ecology of human development. *American Psychologist*, 1977, *32.*

Brunswik, E. Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 1955, *62.*

Campbell, D. T. Degrees of freedom and the case study. *Comparative Political Studies*, 1975, *8.*

Campbell, D. T. Qualitative knowing in action research. *Journal of Social Issues*, in press.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of Research on Teaching*, Chicago, IL: Rand McNally and Company, 1963.

Cicourel, A. V. *Method and measurement in sociology.* New York: NY: Free Press of Glencoe, 1964.

Cronbach, L. J. Course improvement through evaluation. *Teachers College Record*, 1963, *64.*

Cronbach, L. J. Beyond the two disciplines of scientific psychology. *American Psychologist*, 1975, *30.*

Denzin, N. K. The logic of naturalistic inquiry. *Social Forces*, 1971, *50.*

Eisner, E. Models for educational evaluation. *School Review*, 1972, *80.*

Eisner, E. Do behavioral objectives and accountability have a place in art education? *Art Education*, 1973, *26.*

Eisner, E. W. The perceptive eye: Toward the reformation of educational evaluation. *Occasional Papers of the Stanford Evaluation Consortium*, Stanford, CA: Stanford University, 1975, (Mimeo).

Eisner, E. W. *The design and evaluation of educational programs.* New York, NY: The Macmillan Company, in press.

Ellsworth, P. C. From abstract ideas to concrete instances: Some guidelines for choosing natural research settings. *American Psychologist*, 1977, *32*.

Erickson, F. Some approaches to inquiry in school-community ethnography. *Anthropology and Education*, 1977, *8*.

Fienberg, S. E. The collection and analysis of ethnographic data in educational research. *Anthropology and Education*, 1977, *8*.

Glaser, B. G., & Strauss, A. L. *The discovery of grounded theory.* Chicago, IL: Aldine Publishing Company, 1967.

Glass, G. V. A paradox about excellence of schools and the people in them. *Educational Researcher*, 1975, *4*.

Glass, G. V. (Ed.). *Evaluation studies review annual volume I.* Beverly Hills, CA: Sage Publications, 1976.

Goffman, E. *Frame analysis.* New York, NY: Harper Coliphon Books, 1974.

Guba, E. G. *Methodological strategies for educational change.* Paper presented to the Conference on Strategies for Educational Change, Washington, D.C., 1965 (Mimeo).

Guba, E. G. The failure of educational evaluation. *Educational Technology*, 1969, *9*.

Guba, E. G. The state of the art of educational evaluation. *Proceedings of the Evaluation Network Annual Meeting*, St. Louis, MO., 1977, In press.

Guttentag, M. (Ed.). *Evaluation studies review annual, volume 2.* Beverly Hills, CA: Sage Publications, 1977.

Herriott, R. E. Ethnographic case studies in federally funded multidisciplinary policy research: Some design and implementation issues. *Anthropology and Education*, 1977, *8*.

Hersen, M., & Barlow, D. H. *Single case experimental designs.* New York, NY: Pergamon Press, 1976.

House, E. R. (Ed.) *School evaluation: The politics and process.* Berkeley, CA: McCutchan Publishing Corporation, 1973.

House, E. R. Justice in evaluation. In G V Glass (Ed.), *Evaluation Studies Review Annual* (Volume 1). Beverly Hills, CA: Sage Publications, 1976.

House, E. R. *The logic of evaluative argument.* CSE Monograph series in evaluation, No. 7. Los Angeles: Center for the Study of Evaluation, University of California, 1977.

Jackson, P. W. *Life in classrooms.* New York, NY: Holt, Rinehart, and Winston, 1968.

Johnson, J. M. *Doing field research.* New York, NY: The Free Press, 1975.

Joint Committee on Standards for Educational Evaluation. *Standards for educational evaluation.* Kalamazoo, MI: The Committee, Evaluation Center, Western Michigan University, 1977 (Draft, Mimeo).

Kaplan, A. *The conduct of inquiry.* San Francisco, CA: Chandler Publishing Company, 1964.

Kuhn, T. S. The structure of scientific revolutions. *International Encyclopedia of Unified Science, 2nd Edition, 2,* No. 2. Chicago, IL: University of Chicago Press, 1970.

Levine, M. Scientific method and the adversary model. *American Psychologist,* 1974, *29.*

Lutz, F. W., & Ramsey, M. A. The use of anthropological field methods in education. *Educational Researcher,* 1974, *3.*

Malinowski, B. *The argonauts of the western pacific.* London: Routledge and Kegan Paul, 1922.

McCall, G. J., & Simmons, J. L. *Issues in participant observation.* Reading, MA: Addison-Wesley Publishing Company, 1969.

Miller, D. B. Roles of naturalistic observation in comparative psychology. *American Psychologist,* 1977, *32.*

O'Keefe, K. G. *Methodology for educational field studies.* Unpublished Doctoral Dissertation, Ohio State University, 1968.

Owens, T. R. Educational evaluation by adversary proceedings. In E. R. House (Ed.), *School Evaluation: The Politics and Process.* Berkeley, CA: McCutchan Publishing Corporation, 1973.

Parlett, M., & Hamilton, D. Evaluation as illumination: A new approach to the study of innovatory programs. In G V Glass (Ed.), *Evaluation Studies Review Annual* (Volume 1). Beverly Hills, CA: Sage Publications, 1976.

Patton, M. Q. *Alternative evaluation research paradigm.* Study Group on Evaluation, Grand Forks, ND: North Dakota, University of North Dakota, 1975.

Proshansky, H. M. Environmental psychology and the real world. *American Psychologist,* 1976, *31.*

Provus, M. M. *Discrepancy evaluation.* Berkeley, CA: McCutchan Publishing Corporation, 1971.

Rawls, J. *A theory of justice.* Cambridge, MA: Balknop Publishing Company, 1971.

Rippey, R. M. *Studies in transactional evaluation.* Berkeley, CA: McCutchan Publishing Corporation, 1973.

Rist, R. C. Ethnographic techniques and the study of an urban school. *Urban Education,* 1975, *10.*

Rist, R. C. On the relations among educational research paradigms: From disdain to detente. *Anthropology and Education*, 1977, *8*.

Rivlin, A. M. Social experiments: The promise and the problem. *Brookings Bulletin*, 1973, *10*.

Roberts, J. I., & Akinsanya, S. K. *Schooling in the cultural context*. New York, NY: David McKay Publishing Company, 1976.

Rossi, P. H., & Wright, S. R. Evaluation research: An assessment of theory, practice, and politics. *Evaluation Quarterly*, 1977, *1*.

Scriven, M. Objectivity and subjectivity in educational research. In Lawrence G. Thomas (Ed.), *Philosophical Redirection of Educational Research* (71st Yearbook, Part I). Chicago, IL: National Society for the Study of Education, The University of Chicago Press, 1972.

Scriven, M. Maximizing the power of causal investigation: The modus operandi method. In W. J. Popham (Ed.), *Evaluation in Education*. Berkeley, CA: McCutchan Publishing Corporation, 1974.

Semmel, M. I., & Thiagarajan S. Observation systems and the special education teacher. *Focus on Exceptional Children*, 1973, *5*.

Shrock, S. Ecological inquiry: *Expanding the definition*. Paper presented at the Annual Meeting of the Association for Educational Communications and Technology, Miami, 1977.

Smith, E. R., & Tyler, R. W. *Appraising and recording student progress*. New York, NY: Harper, 1942.

Speizman, W. *The purposes of exploratory field work*. Unpublished document, n.d. (Mimeo).

Stake, R. E. The countenance of educational evaluation. *Teachers College Record*, 1967, *68*.

Stake, R. E. *Evaluating the arts in education*. Columbus, OH: Charles E. Merrill Publishing Company, 1975a.

Stake, R. E. Program evaluation, particularly responsive evaluation (Occasional Paper No. 5) The Evaluation Center, Western Michigan University, 1975b.

Stake, R. E. A review of the review of the handbook. *Journal of Educational Measurement*, 1977, *14*.

Stanford Evaluation Consortium, (Cronbach, J., & Ross, L. Eds.). Handbook of evaluation research. *Educational Researcher*, 1976, *5*.

Stufflebeam, D. L., et al. *Educational evaluation and decision-making*. Itasca, IL: Peacock Publishing Company, Inc., 1971.

Stufflebeam, D. L. *Administrative checklist for reviewing evaluation plans*. Evaluation Center, Western Michigan University, 1974 (Mimeo).

Talmadge, H. Evaluation of local school community problems: A transactional evaluation approach. *Journal of Research and Development in Education*, 1975, *8*.

Tyler, R. W. *Syllabus for education 360.* Chicago, IL: University of Chicago Press, 1960.

Tymitz, B., & Wolf, R. *An introduction to judicial evaluation and natural inquiry.* Nero and Associates, 1977 (Mimeo).

Webb, E. J., et al. *Unobtrusive measures.* Chicago, IL: Rand McNally and Company, 1966.

Weiss, C. H. (Ed.). *Evaluating action programs.* Boston, MA: Allyn and Bacon, 1972.

Willems, E. P., & Raush, H. L. *Naturalistic viewpoints in psychological research.* New York, NY: Holt, Rinehart, and Winston, 1969.

Wilson, S. The use of ethnographic techniques in educational research. *Review of Educational Research*, 1977, *47.*

Wolcott, H. *The man in the principal's office: An ethnography.* New York, NY: Holt, Rinehart, and Winston, 1973.

Wolf, R. L. *The application of select legal concepts in educational evaluation.* Unpublished Doctoral Dissertation, University of Illinois, 1974.

Wolf, R. L. Trial by jury: A new evaluation method. *Phi Delta Kappan*, 1975, *57.*

Wolf, R. L., & Tymitz, B. Ethnography and reading; Matching inquiry mode to process. *Reading Research Quarterly*, 1976-77, *12.*

Wolf, R. L., & Tymitz, B. Toward more natural inquiry in education. *CEDR Quarterly*, 1977, *10.*

Worthen, B. R. Characteristics of good evaluation studies. *Journal of Research and Development in Education*, 1977, 10.